

FOR OFFICIAL USE ONLY

ACCESS DB # _____

PLEASE PRINT CLEARLY

Location (Bldg/Room#): _____

Scientific and Technical Information Center

SEARCH REQUEST FORM

Date: _____ Requester's Full Name: _____ Examiner #: _____

Art Unit: _____ Phone (301) _____ Serial Number: _____

Results Format Preferred (circle): PAPER DISK E-MAIL

To ensure an efficient and quality search, please attach a copy of the cover sheet, claims, and abstract or fill out the following:

Title of Invention: _____

Inventors (please provide full names): _____

Earliest Priority Date: _____

Search Topic:

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known.

For Sequence Searches Only Please include all pertinent information (parent, grandchild, divisional, or issued patent numbers) along with the appropriate serial number.

STAFF USE ONLY

Searcher: Beverly C 4999 Type of Search _____ NA Sequence (#)

Searcher Phone #: _____ AA Sequence (#)

Searcher Location: _____ Structure (#)

Date Searcher Picked Up: _____ Bibliographic

Date Completed: 02-18-00 Litigation

Searcher Prep & Review Time _____ Fulltext

Online Time: _____ Other

Vendors and Cost

_____ STN _____ Dialog

_____ Questel/Orbit _____ Dr. Link

_____ Lexis/Nexis _____ Westlaw

_____ WWW/Internet

_____ In-house sequence systems (list)

_____ Other (specify) MP

THIS PAGE BLANK (USPTO)

CC modulating the production of a T-helper type 2 associated cytokine.
CC Especially the method is useful in the inhibition of Th2 in allergy/
CC cancer or infections, and promotion of Th2 in autoimmune disease and
CC transplantation.
SQ Sequence 412 AA;

Sequence 412 AA;

Query Match	100.0%;	Score 2858;	DB 28;	Length 412;
Best Local Similarity	100.0%;	Pred. No. 2.68e-219;		
Matches 412;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

Dd 1 maepirgrgsprsgyrgarrarargrgrcparqsparlipdtvlvdlvsdsdeevleavd 600
Qy 1 MAEPLRGGRPSRSGRGARARRARGRCPRARQSPARLPDITVLVDLVSDDSDEEVLEAVD 600

Db 61 pvevparvarlpapakpeqdsdsassegaaegpagaprtlrvrrrrrrllldpgaeppvpyvsqk 120
 |||||
 61 pvevparvarlpapakpeqdsdsdssegaaegpagaprtlrvrrrrrrrrllldpgaeppvpyvsqk 120

121 vqsslnllpnssllklcpsepedeadi.tnsgsspsedda.lpsgprwkr.lrkckceek 180
+++++
121 vqsslnllpnssllklcpsepedeadi.tnsgsspsedda.lpsgprwkr.lrkckceek 180
+++++

```
D6      181 kmeepdqdspipqpsrnkrrkhteaqlkrevnkrigdirscispkgqspalqstd 2400
OY      181 kmeepddqisplppqssrnksrkhtelqlkrevnkrldlrscslspkhohspaldstd 2400
```

D6 241 deavlvpgvlpqpsrilitkircradlvr lprvmseplgnvcdmahnlgvspnrlll 3000
|||||
|||
QY 241 DEAVLVGPVLPOSSRLFTLKIRCRA DLVLRPVPMSEPLGNVDHMHNLGVSPNRILL 3000

Db 301 tgeselptatpstkrgvadidcvtlasssestetqelrtrvgkexqmleislp 360
 |||||
 2y 301 fgeelstptatpstkrgvadidcvtlasssestetqelrtrvgkexqmleislp 360

D5 361 dsprkvlmsrhyeeanglsngkrstirigcklssgelpadglsgdlievwg 412
 |||||
Q7 361 dsprkvlmsrhyeeanglsngkrstirigcklssgelpadglsgdlievwg 412

RESULT	2
ID	W60079 standard; Protein; 101 AA.
AC	W60079:

DT 11-SEP-1998 (first entry)
DE Homo sapiens sentrin-1 polypeptide.
FN sentrin-1; protection; tumour necrosis factor; tnfr; apoptosis;
Fas/APO-1-induced: tumour cell death: induction; tumour aggressiveness:

CS detection; determination.
 PN Homo sapiens.
 PD WO9820038-A1.
 14-MAY-1998

PI Yeh ETH:
PA (TEXA) UNIV TEXAS SYSTEM.
PR 05-NOV-1996; US-030302.
PR 03-NOV-1997; 020344.
PR

PT New isolated sentrin polypeptide(s) - which inhibit TNF receptor or Fas/APO-induced apoptosis, used to develop products for inducing DR WPI: 98-266868/25.
DR N-PSDB: V34564.

PI cell death in tumours
PS Claim 4; Page 78; 120pp; English.
CC The sequence is that of the sentrin-1 polypeptide.
CC Sentrin polypeptides have the ability to protect or guard

cells from tumour necrosis factor (TNF) or Fas/APO induced cell death (apoptosis). Inhibitors of the sentrin polypeptides, e.g. antibodies, can be used for inducing cell death, particularly in tumours. The products can also be used for

Sequence 101 AA: detect a ubiquitin conjugating enzyme polypeptide or PM1 polypeptide.

Query Match	5.38;	Score 151;	DB 32;	Length 101;
Best Local Similarity	27.48;	Pred. No. 3.49e-02;		

Matches 23; Conservative 27; Mismatches 31; Indels 3; Gaps 3;

```
Dd      8 patedlgdkkkgseyiklvigqdsael-hfkvkmthtlklkesyccrgypmmslflf 66
       ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy    329 ASSESEATEISQ-E-LRLRVGGEKHQLLEISLSPDSPLVLMSHYEAMGLSGHKLSFFF 386
```

```

Db      67  eggrladnhpkelgmeedvlev 90
      :|::: |::|:| |::|
QY     387 DGTKSGKELPADLGESGDLIEV 410

```

RESULT	3
ID	W87497 standard; Protein; 126 AA.
NC	W87497.

DT 18-FEB-1999 (first entry)
DE Human SMT3-like protein (HSMTH).
KW HSMTH; SMT3-like protein; recombinant; foetal development; inflammation;
cancer; radiation; DNA damage; carcinogenesis; stress; ethanol; bacteria;

KW sunscren; therapeutic; cell proliferative disease; antagonist; asthma;
 Crohn's disease; rheumatoid arthritis; ribozyme; chromosome mapping;
 immunoassay; drug; screening.
 KW Homo sapiens

	Location/Qualifiers
FH Key	12
FT Modified_site	/note= "potential phosphorylation site"
FT Modified_site	23

FT	/note="potential myristoylation site"
FT	34
FT	Modified_site
FT	Modified_site
FT	53
FT	/note="potential phosphorylation site"

FT	/note="potential phosphorylation site"
FT	63
Modified_site	
FT	/note="potential myristoylation site"
FT	67
Modified_site	

FT	/note= "potential N-glycosylation site"
FT	88
FT	Modified_site
FT	/note= "potential phosphorylation site"
FT	108
FT	Misc difference

FT	/label= unknown
FT	/note= "encoded by ANA"
PN	
DD	12-NOV-1998
	W09850545-A1.

PF 06-MAY-1998; 008420.
PR 09-MAY-1997; US-853974.
PA (INCY-) INCYTE PHARM INC.
Hillman II, Shah P.
I

DR MPI: 99-034/20/03.
DR N-PSDB; V79504.
PT New human SKI3-like protein - for treatment, prevention and diagnosis of cancer and inflammation, and for promoting DNA repair

CC This represents a human SMT3-like protein (HSMTH). Host cells transformant
CC with expression vectors containing the HSMTH nucleic acid are used for
CC the recombinant production of the protein. HSMTH, expressed in cDNA
PS Claim 1, Fig. 1A-B; 55pp; English.

CC radiation damage, is used to promote DNA repair (before, during or after
CC exposure to damage from radiation or carcinogenic compounds). The
CC polyepoxide can be used to treat patients with ataxia telangiectasia. On

CC associated with these diseases, or it is added to sunscreens. The HSMTH
CC protein is also used for the generation of antibodies and screening for
CC specific binding agents (potential therapeutic). Antisense agents are used to
CC related diseases, who are undergoing radiation treatment for cancers.

CC treat or prevent cell proliferative diseases, especially cancer and
CC inflammation (e.g. asthma, Crohn's disease, rheumatoid arthritis). The
CC HSM11 nucleic acid and its fragments are used as antisense/ribozyme
CC therapeutics: for detecting and quantifying gene expression; for

CC isolating related sequences and for chromosome mapping. The antibodies
CC are used directly as therapeutic antagonists; for delivering drugs to
CC cells expressing HSMTH; as immunoassay reagents for diagnosis or
CC monitoring disease or treatment; in competitive drug screens and to

CC Isolate the protein from natural sources.
50 Sequence 126 AA;

Query Match	Similarity	Score	DB	Length
Best Local	31.6%	Pred. No. 2,99e-02;		
Matches	25;	Conservative	19;	Mismatches 33; Indels 2; Gaps 2
Db	8	evgkxendh1nk1kvaggds-vvqf1k1rhtplek1mkaycerg1asmrqtfrfdgqpl	66	
Oy	333	ENTERSEQ-LILNRQGRKKHOMLEISLSPDPLAVLMSHYEAMGLSGHKLSTFFDQTKL	391	
Db	67	netdtpaqlemededtlidv	85	
Oy	392	SGKEPLADLGESGDLIEV	410	
RESULT	4			
ID	W74776	standard; Protein; 102 AA.		
AC	W74776;			
DT	25-JAN-1999	(first entry)		
DE	Human secreted protein encoded by gene 47 clone HOGAV75			
KW	Human; secreted protein; testis; tumour; foetal brain tissue;			
FX	Fusion protein; cancer; central nervous system; seizure;			
DI	diagnosis; neurodegenerative disease.			
OS	Homo sapiens.			
FT	Key	Location/Qualifiers		
FT	Misc-difference	102		
FT		label= unknown		
PN	W09839448-A2.			
PD	11-SEP-1998.			
PE	06-MAR-1998;	U04493.		
PF	02-OCT-1997;	US-061060.		
PR	07-MAR-1997;	US-038621.		
PR	07-MAR-1997;	US-040161.		
PR	07-MAR-1997;	US-040162.		
PR	07-MAR-1997;	US-040163.		
PR	07-MAR-1997;	US-040333.		
PR	07-MAR-1997;	US-040334.		
PR	07-MAR-1997;	US-040336.		
PR	07-MAR-1997;	US-040626.		
PR	11-APR-1997;	US-043311.		
PR	11-APR-1997;	US-043312.		
PR	11-APR-1997;	US-043313.		
PR	11-APR-1997;	US-043314.		
PR	11-APR-1997;	US-043568.		
PR	11-APR-1997;	US-043569.		
PR	11-APR-1997;	US-043576.		
PR	11-APR-1997;	US-043578.		
PR	11-APR-1997;	US-043580.		
PR	11-APR-1997;	US-043659.		
PR	11-APR-1997;	US-043670.		
PR	11-APR-1997;	US-043671.		
PR	11-APR-1997;	US-043672.		
PR	11-APR-1997;	US-043674.		
PR	23-MAY-1997;	US-047492.		
PR	23-MAY-1997;	US-047500.		
PR	23-MAY-1997;	US-047501.		
PR	23-MAY-1997;	US-047502.		
PR	23-MAY-1997;	US-047503.		
PR	23-MAY-1997;	US-047581.		
PR	23-MAY-1997;	US-047582.		
PR	23-MAY-1997;	US-047583.		
PR	23-MAY-1997;	US-047584.		
PR	23-MAY-1997;	US-047585.		
PR	23-MAY-1997;	US-047586.		
PR	23-MAY-1997;	US-047587.		
PR	23-MAY-1997;	US-047588.		
PR	23-MAY-1997;	US-047589.		
PR	23-MAY-1997;	US-047590.		
PR	23-MAY-1997;	US-047592.		
PR	23-MAY-1997;	US-047593.		
PR	23-MAY-1997;	US-047594.		
PR	23-MAY-1997;	US-047595.		
PR	23-MAY-1997;	US-047596.		
PR	23-MAY-1997;	US-047597.		
PR	23-MAY-1997;	US-047598.		

23-MAY-1997; US-047599.
PR 23-MAY-1997; US-047600.
PR 23-MAY-1997; US-047601.
PR 23-MAY-1997; US-047612.
PR 23-MAY-1997; US-047613.
PR 23-MAY-1997; US-047614.
PR 23-MAY-1997; US-047615.
PR 23-MAY-1997; US-047617.
PR 23-MAY-1997; US-047618.
PR 23-MAY-1997; US-047632.
PR 23-MAY-1997; US-047633.
PR 06-JUN-1997; US-048964.
PR 06-JUN-1997; US-048974.
PR 13-JUN-1997; US-049610.
PR 08-JUL-1997; US-051926.
PR 16-JUL-1997; US-052874.
PR 18-AUG-1997; US-055724.
PR 22-AUG-1997; US-056630.
PR 22-AUG-1997; US-056631.
PR 22-AUG-1997; US-056632.
PR 22-AUG-1997; US-056636.
PR 22-AUG-1997; US-056637.
PR 22-AUG-1997; US-056662.
PR 22-AUG-1997; US-056664.
PR 22-AUG-1997; US-056845.
PR 22-AUG-1997; US-056846.
PR 22-AUG-1997; US-056862.
PR 22-AUG-1997; US-056864.
PR 22-AUG-1997; US-056872.
PR 22-AUG-1997; US-056874.
PR 22-AUG-1997; US-056875.
PR 22-AUG-1997; US-056876.
PR 22-AUG-1997; US-056877.
PR 22-AUG-1997; US-056878.
PR 22-AUG-1997; US-056879.
PR 22-AUG-1997; US-056880.
PR 22-AUG-1997; US-056881.
PR 22-AUG-1997; US-056882.
PR 22-AUG-1997; US-056884.
PR 22-AUG-1997; US-056886.
PR 22-AUG-1997; US-056887.
PR 22-AUG-1997; US-056888.
PR 22-AUG-1997; US-056889.
PR 22-AUG-1997; US-056892.
PR 22-AUG-1997; US-056893.
PR 22-AUG-1997; US-056894.
PR 22-AUG-1997; US-056903.
PR 22-AUG-1997; US-056908.
PR 22-AUG-1997; US-056909.
PR 22-AUG-1997; US-056910.
PR 22-AUG-1997; US-056911.
PR 05-SEP-1997; US-057650.
PR 05-SEP-1997; US-057669.
PR 05-SEP-1997; US-057761.
PR 12-SEP-1997; US-058785.
PR 12-SEP-1997; US-058785.
PA (HUMA-) HUMAN GENOME SCI. INC.
PI Bednarik DP, Brewer LA, Carter KC, Duan R, Endress GA,
PI Feng P, Ferlie AP, Fischer CL, Florence KA, Greene JM, Hu JS,
PI Kyaw H, Lafleur DW, Li Y, Moore PA, Ni J, Olsen HS, Rosen CA,
PI Ruken SM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z.
PI WPI: 98-506364/43.
DR N-PSDB: V59557.
PR New isolated human genes and the secreted polypeptide(s) they encode
PT - useful for diagnosis and treatment of e.g. cancers, neurological
PT disorders, immune diseases, inflammation or blood disorders
PS Claim 1; Page 566-567; 721pp; English.
CC This sequence represents a secreted human protein encoded by the nucleic
CC acid molecule designated Gene 47 from the human cDNA clone HOGANV75
CC (deposited as clone ATCC 97899 and ATCC 209045).
CC The gene can be used to generate fusion proteins by linking to the gene
CC to a human immunoglobulin Fc portion (e.g. V59502) for increasing the
CC stability of the fused protein as compared to the human protein only.
CC The invention relates to 186 novel genes and their fragments (nucleic
CC acid sequences: V59511-V59812; amino acid sequences W74731-W75026) which


```

CC   cyclase
SO   Sequence 1184 AA:

Query Match
Best Local Similarity 3.8%; Score 109; DB 14; Length 1184;
Matches 19; Conservative 9; Mismatches 8; Indels 2; Gaps 2

      84  applggagprraagppgrrarargrrrprprrgagrr 121
      ||| ||| : ||| ||| : ||| ||| : ||| ||| : |||
      2  AEPPLRGPRRSRGGRARRARG-ARGRCPRAR-OSPAR 37

RESULT 8
ID   R32882 standard; Protein; 1184 AA.
AC   R32882.
DT   17-JUN-1993 (first entry)
DE   Cardiac adenyl cyclase type V.
KW   CACV; therapy; diagnostic; cardiac function; cyclic AMP; cAMP; heart
KW   failure.
OS   Canis familiaris.
EP   E29622-A.
PR   03-MAR-1993.
PR   27-AUG-1992; 114637.
PA   (AMCY ) AMERICAN CYANAMID CO.
PI   Ishikawa Y, Kanski AF;
PI   WPI: 93-06868/09.
DR   N-PSDB; 037543.
PT   Isolated nucleic acid mol. encoding Cardiac adenyl cyclase type
PT   V - useful for determining and modifying cardiac function
PS   Claim 4; Page 15-27; 38pp. English.
CC   Left ventricular tissue of canine heart was used as a source of mRNA.
CC   A cDNA library was prepd. In lambda gt10 phage. A 970 bp Aat-HincII
CC   fragment from type I adenyl cyclase cDNA was used as probe. The
CC   clones isolated were used to obtain cDNA encoding CACV. This probe
CC   may also be used to screen a human cardiac cDNA library to obtain
CC   the cDNA encoding human CACV. CACV, its analogues and antibodies
CC   are useful in therapy or diagnostic assays, e.g. in modifying and
CC   determining cardiac function. A decrease in CACV content of the
CC   heart contributes to impaired cAMP prodn. and in heart failure. The
CC   CACV can also be used to screen for cpds. which stimulate or inhibit
CC   the activity of the cyclase.
CC   Sequence 1184 AA:

Query Match
Best Local Similarity 3.8%; Score 109; DB 6; Length 1184;
Matches 19; Conservative 9; Mismatches 8; Indels 2; Gaps 2

      84  applggagprraagppgrrarargrrrprprrgagrr 121
      ||| ||| : ||| ||| : ||| ||| : ||| ||| : |||
      2  AEPPLRGPRRSRGGRARRARG-ARGRCPRAR-OSPAR 37

RESULT 9
ID   W68479 standard; Protein; 122 AA.
AC   W68479.
DT   08-DEC-1998 (first entry)
DE   HIV-1 strain YBF30 rev protein.
KW   HIV-1 strain YBF30; antibody; oligonucleotide; diagnosis; immunisation;
KW   infection; typing; rev.
OS   Human immunodeficiency virus type 1.
FT   key
FT   Location/Qualifiers
FT   MISC_difference 104
FT   /label= unknown
FT   /note= "encoded by TGA"
PD   12-JUN-1998.
PF   09-DEC-1996; 015087.
PF   09-DEC-1996; FR-015087.
PA   (ASSI-) ASSISTANCE PUBLIQUE HOPITAUX PARIS.
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
PA   (INSP ) INST PASTEUR.
PI   Baire-Sinoussi F, Lousserat-Ajaka I, Mauciere P, Saragosti S,

```

[illegible]

THIS PAGE BLANK (USPTO)

(TM)

Distribution rights by Oxford Molecular Ltd

41	94	1618	1	US-07-853-3	Applicatio	7.28e+01
42	92	222	2	US-08-993-3	Applicati	9.53e+01
43	92	351	1	US-08-402-2	Applicatio	9.53e+01
44	92	704	1	US-08-646-6	Sequence 5,	9.53e+01
45	92	1612	3	Sequence 48,	Applicati	9.35e+01
				PCT-US94-0		

ALIGNMENTS

RESULT 1

RESULT	1	STANDARD;	PRT;	412 AA
ID	US-08-755-584-2			
XX				
AC	xxxxxx			

RESULT	1	STANDARD;	PRT;	412 AA
ID	US-08-755-584-2			
XX				
AC	xxxxxx			
XX				
DT				

RESULT	1	STANDARD;	PRT;	412 AA
ID	US-08-755-584-2			
XX				
XX	xxxxxx			
AC				
XX				
DT				
XX				
DE	Sequence 2, Application US/08755584			
XX				

	RESULT	1			
ID	US-08-755-584-2		STANDARD;	PRT;	412 AA.
XX					
AC	xxxxxx				
XX					
DJ					
XX					
DE	Sequence 2, Application US/08755584				
CC	Sequence 2, Application US/08755584				
CC	Patent No. 5658711				
CC	GENERAL INFORMATION:				

RESULT	1			
ID	US-08-755-584-2	STANDARD;	PRT;	412 AA.
XX				
AC	xxxxxx			
XX				
DT				
XX				
DE				
XX				
	Sequence 2, Application US/08755584			
CC				
CC	Sequence 2, Application US/08755584			
CC	Patent No. 5858711			
CC	GENERAL INFORMATION:			
CC	APPLICANT: Glimcher, Laurie H.			
CC	APPLICANT: Hodge, Martin R.			

RESULT	1	STANDARD;	PRT;	412	AA.
ID	US-08-755-584-2				
XX					
XX	xxxxxx				
AC					
XX					
DT					
XX					
DE	Sequence 2, Application US/08755584				
XX					
CC	Sequence 2, Application US/08755584				
CC	Patent No. 5858711				
CC	GENERAL INFORMATION:				
CC	APPLICANT: Glimcher, Laurie H.				
CC	APPLICANT: Hodges, Martin R.				
CC	TITLE OF INVENTION: NF-AT-INTERACTING PROTEIN NIP45 AND METHODS				
CC	TITLE OF INVENTION: OF USE THEREFOR				
CC	NUMBER OF SEQUENCES: 2				
CC	CORRESPONDENCE ADDRESS:				

RESULT	1	STANDARD;	PRT;	412	AA.
ID	US-08-755-584-2				
XX					
XX	xxxxxx				
XX					
DT					
XX					
DE	Sequence 2, Application US/08755584				
XX					
CC	Sequence 2, Application US/08755584				
CC	Patent No. 5858711				
CC	GENERAL INFORMATION:				
CC	APPLICANT: Glimecher, Laurie H.				
CC	ATTORNEY: Hodge, Martin R.				
CC	TITLE OF INVENTION: NF-AT-INTERACTING PROTEIN NIP4S AND METHODS				
CC	TITLE OF INVENTION: OF USE THEREFOR				
CC	NUMBER OF SEQUENCES: 2				
CC	CORRESPONDENCE ADDRESS:				
CC	ADDRESSEE: LAHIVE & COCKFIELD				
CC	STREET: 60 State Street, suite 510				
CC					

RESULT	1	STANDARD;	PRT;	412 AA.
ID	US-08-755-584-2			
XX				
AC	xxxxxx			
XX				
DT				
XX				
DE	Sequence 2, Application US/08755584			
XX				
CC	Sequence 2, Application US/08755584			
CC	Patent No. 5658711			
CC	GENERAL INFORMATION:			
CC	APPLICANT: Glimcher, Laurie H.			
CC	APPLICANT: Hodge, Martin R.			
CC	TITLE OF INVENTION: NF-AT-INTERACTING PROTEIN NIP45 AND METHODS			
CC	TITLE OF INVENTION: OF USE THEREFOR			
CC	NUMBER OF SEQUENCES: 2			
CC	CORRESPONDENCE ADDRESS:			
CC	ADDRESSEE: LAHIVE & COCKFIELD			
CC	STREET: 60 State Street, suite 510			
CC	CITY: Boston			
CC	STATE: Massachusetts			
CC	COUNTRY: USA			
CC	ZIP: 02109-1875			
CC	COMPUTER READABLE FORM:			
CC	MEDIUM TYPE: FLOPPY disk			
CC	COMPUTER: IBM PC compatible			
CC	OPERATING SYSTEM: PC-DOS/MS-DOS			
CC	SOFTWARE: Patentin Release #1.0, Version #1.25			
CC	CURRENT APPLICATION DATA:			
CC	APPLICATION NUMBER: US/08/755,584			
CC	FILING DATE:			
CC	CLASSIFICATION: 435			
CC	PRIOR APPLICATION DATA:			
CC	APPLICATION NUMBER:			
CC	FILING DATE:			
CC	ATTORNEY/AGENT INFORMATION:			
CC	NAME: Giulio A. DeConti, Jr.			
CC	REGISTRATION NUMBER: 31,503			
CC	REFERENCE/DOCKET NUMBER: HUT-026			
CC	TELECOMMUNICATION INFORMATION:			
CC	TELEPHONE: (617)227-7400			
CC	TELEFAX: (617)227-5941			
CC	INFORMATION FOR SEQ ID NO: 2:			
CC	SEQUENCE CHARACTERISTICS:			
CC	LENGTH: 412 amino acids			
CC	TYPE: amino acid			


```

CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC IMMEDIATE SOURCE:
CC LIBRARY: LUNGFET03
CC CLONE: 2361410
CC SEQUENCE 126 AA; 13961 MW; 87522 CN;

Query Match 5.3%; Score 152; DB 2; Length 126;
Best Local Similarity 31.6%; Pred. No. 1,57e-02;
Matches 25; Conservative 19; Mismatches 33; Indels 2; Gaps 2;

Db 8 EGVKENDHINKVAGDGS-VVQFKIRKHPPLSLMKAYCROGLSMROIREFRDGQPI 66
OY 333 EATETTSOE-LRLRVGKKEHOMLEISLSPDSLKAVLMSHYEAMGLSGHKLSFFPDGKTL 391
OY 67 NETDTPAOLEMEDEDITDV 85
OY 392 SGKELPADLGLESGDLIEV 410

RESULT 4
US-08-455-073A-6 STANDARD; PRT; 673 AA.
xxxxxx

Sequence 6, Application US/08455073A
Sequence 6, Application US/08455073A
Patent No. 5876949
GENERAL INFORMATION:
APPLICANT: Gideon Dreyfuss
APPLICANT: Mikiko C. Sioml
APPLICANT: Yan Zhang
TITLE OF INVENTION: Fragile X Related Proteins, Compositions And Methods
TITLE OF INVENTION: Of Making And Using The Same
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & NO. 5876949rls
STREET: One Liberty Place, 46th floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Wordperfect 6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/455,073A
FILING DATE: 31-MAY-1995
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: DeLuca, Mark
REGISTRATION NUMBER: 33,229
REFERENCE/DOCKET NUMBER: UPN-2201
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3410
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 673 amino acid residues
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
SEQUENCE 673 AA; 74079 MW; 2221704 CN;

Query Match 4.0%; Score 113; DB 2; Length 673;
Best Local Similarity 32.4%; Pred. No. 5.18e+00;
Matches 23; Conservative 18; Mismatches 26; Indels 4; Gaps 4;

```

```

Db      471 RDRTGCESSRRRTGGGRPPAPAPTSTSNSSSTISVYLKDPDSDPYSLDTSP-EP 529
Oy      8 RGPSPRGGRGRARRRARGAR-CPRARSPALIPDTIVLVLY-SDSDE-EYLEVADEVY 64
Db      530 PVDSPEGEPPP 540
Oy      65 PVARLPAPAKP 75
RESULT          5              STANDARD:      PRT;      265 AA.
ID US-08-970-133-1
XX AC xxxxxx
XX DT
XX DE
XX SE Sequence 1, Application US/08970133
CC CC Sequence 1, Application US/08970133
CC Patent No. 5916753
CC GENERAL INFORMATION:
CC APPLICANT: Bandman, Olga
CC APPLICANT: Guegler, Karl J.
CC APPLICANT: Lai, Preeli
CC TITLE OF INVENTION: SH3-CONTAINING PROTEINS
CC NUMBER OF SEQUENCES: 5
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Incyte Pharmaceuticals, Inc.
CC STREET: 3174 Porter Dr.
CC CITY: Palo Alto
CC STATE: CA
CC COUNTRY: USA
CC ZIP: 94304
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette
CC COMPUTER: IBM Compatible
CC OPERATING SYSTEM: DOS
CC SOFTWARE: Fastspo for Windows Version 2.0
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/970,133
CC FILING DATE: Filed Herewith
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER:
CC FILING DATE:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Billings, Lucy J.
CC REGISTRATION NUMBER: 36,749
CC REFERENCE/DOCKET NUMBER: PF-0419 US
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 650-855-0555
CC TELEFAX: 650-845-4166
CC INFORMATION FOR SEQ ID NO: 1:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 265 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC IMMEDIATE SOURCE:
CC LIBRARY: BRAITUT03
CC CLONE: 865744
SQ SEQUENCE 265 AA; 30106 MW; 347009 CN;
Query Match 3.7%; Score 106; DB 2; Length 265;
Best Local Similarity 28.6%; Pred. No. 1,3se+01;
Matches 22; Conservative 21; Mismatches 30; Indels 4; Gaps 4.
Db      38 NKGATPEDFSNLPE-QRRKKLQGVDELNKEIQEKENDORDAIFKMVDYLKNPWGDPA 96
Oy      150 NSSGSPEDDALPGSGWRKKLRKCKEKKXME-EFPDD-I SPLQPSSRNKRKHT 207
Db      97 SLDRKLAEVSNNIEKR 113

```

QY 208 AL-OKLEVNKRLODLR 223

RESULT 6
ID PCT-US95-16216-1 STANDARD; PRT: 3248 AA.

XX AC xxxxxx

DE Sequence 1, Application PC/TUS9516216

XX Sequence 1, Application PC/TUS9516216

CC GENERAL INFORMATION:

CC APPLICANT: Yen, Timothy J.

CC APPLICANT: Rattner, Jerome B.

CC TITLE OF INVENTION: Nucleic Acid Encoding a Transiently

CC TITLE OF INVENTION: Expressed Kinetochores Protein, and Methods of Use

CC NUMBER OF SEQUENCES: 4

CC CORRESPONDENCE ADDRESS:

CC ADDRESSEE: Dann, Dorfman, Herrell and Skillman

CC STREET: 1601 Market Street Suite 720

CC CITY: Philadelphia

CC STATE: PA

CC COUNTRY: USA

CC ZIP: 19103-2307

CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Floppy disk

CC COMPUTER: IBM PC compatible

CC OPERATING SYSTEM: PC-DOS/MS-DOS

CC SOFTWARE: Patentin Release #1.0, Version #1.30

CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: PCT/US95/16216

CC FILING DATE:

CC CLASSIFICATION:

CC PRIORITY APPLICATION DATA:

CC APPLICATION NUMBER: US 08/353,700

CC FILING DATE: 09-DEC-1995

CC ATTORNEY/AGENT INFORMATION:

CC NAME: Reed, Janet E.

CC REGISTRATION NUMBER: 36,252

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: (215) 563-4100

CC TELEFAX: (215) 563-4044

CC INFORMATION FOR SEQ ID NO: 1:

CC LENGTH: 3248 amino acids

CC TYPE: amino acid

CC STRANDEDNESS: not relevant

CC TOPOLOGY: not relevant

CC MOLECULE TYPE: protein

CC HYPOTHETICAL: NO

CC ANTI-SENSE: NO

CC SEQUENCE 3248 AA: 372207 MW: 51689535 CN:

SQ

Query Match 3.7%; Score 107; DB 3; Length 3248;

Best Local Similarity 26.5%; Pred. No. 1.21e+01;

Matches 26; Conservative 23; Mismatches 44; Indels 5; Gaps 5;

DB 31 KKEKQOQFOLDISLEAPQOKOTQVENEKTEG-TNLRKRNORLMEICESLEKTKOKISHE 89

QY 177 KEEKMEEFPPDQDISPLPQSSSRKSRKHTALOKLREVNKRLODLRSCLS-PROHOSPA 235

DB 90 LQVNESQVNFQOLNSGKROIEKLEQELKRCSELER 127

QY 236 LQSTDEVVLVEGPLYVLPSSRLFTL-K-I-RCRADLYR 270

RESULT 7

ID US-08-353-700-1 STANDARD; PRT: 3248 AA.

XX AC xxxxxx

DE

DT Sequence 1, Application US/08353700

XX Sequence 1, Application US/08353700

CC Patent No. 5599919

CC GENERAL INFORMATION:

CC APPLICANT: YEN, TIMOTHY J.

CC APPLICANT: RATTNER, JEROME B.

CC TITLE OF INVENTION: NUCLEIC ACID ENCODING A

CC TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,

CC TITLE OF INVENTION: AND METHODS OF USE

CC NUMBER OF SEQUENCES: 4

CC CORRESPONDENCE ADDRESS:

CC ADDRESSEE: DANN, DORFMAN, HERRELL AND SKILLMAN

CC STREET: 1601 MARKET STREET, SUITE 720

CC CITY: PHILADELPHIA

CC STATE: PA

CC COUNTRY: USA

CC ZIP: 19103-2307

CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Floppy disk

CC COMPUTER: IBM PC compatible

CC OPERATING SYSTEM: PC-DOS/MS-DOS

CC SOFTWARE: Patentin Release #1.0, Version #1.25

CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: US/08/353,700

CC FILING DATE: 09-DEC-1994

CC CLASSIFICATION: 435

CC ATTORNEY/AGENT INFORMATION:

CC NAME: REED, JANET E.

CC REGISTRATION NUMBER: 36,252

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: (215) 563-4100

CC TELEFAX: (215) 563-4044

CC INFORMATION FOR SEQ ID NO: 1:

CC LENGTH: 3248 amino acids

CC TYPE: amino acid

CC STRANDEDNESS: single

CC TOPOLOGY: linear

CC MOLECULE TYPE: protein

CC HYPOTHETICAL: NO

CC ANTI-SENSE: NO

CC ORIGINAL SOURCE:

CC ORGANISM: HUMAN

SQ SEQUENCE 3248 AA: 372207 MW: 51689535 CN;

Query Match 3.7%; Score 107; DB 1; Length 3248;

Best Local Similarity 26.5%; Pred. No. 1.21e+01;

Matches 26; Conservative 23; Mismatches 44; Indels 5; Gaps 5;

DB 31 KKEKQOQFOLDISLEAPQOKOTQVENEKTEG-TNLRKRNORLMEICESLEKTKOKISHE 89

QY 177 KEEKMEEFPPDQDISPLPQSSSRKSRKHTALOKLREVNKRLODLRSCLS-PROHOSPA 235

DB 90 LQVNESQVNFQOLNSGKROIEKLEQELKRCSELER 127

QY 236 LQSTDEVVLVEGPLYVLPSSRLFTL-K-I-RCRADLYR 270

RESULT 8

ID US-08-093-453B-2 STANDARD; PRT: 2205 AA.

XX AC xxxxxx

DE

Sequence 2, Application US/08093453B

CC Sequence 2, Application US/08093453B

CC Patent No. 5439816

CC GENERAL INFORMATION:

DE

CC APPLICANT: Frey, Terry K.
CC APPLICANT: Dominguez, Geraldina
CC APPLICANT: Wang, Chin Yen
CC TITLE OF INVENTION: Modified Infectious Rubella Virus
CC NUMBER OF SEQUENCES: 13
CC CORRESPONDENCE ADDRESSES:
CC ADDRESSEE: Jamie L. Greene, Jones & Askew
CC STREET: 191 Peachtree Street, 37th Floor
CC CITY: Atlanta
CC STATE: Georgia
CC COUNTRY: United States
CC ZIP: 30303
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette, 3.50
CC COMPUTER: Macintosh
CC OPERATING SYSTEM: 7.0
CC SOFTWARE: Microsoft Word
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/093.453B
CC FILING DATE: 19 JUL 1993
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: U.S. 07/722,334
CC FILING DATE: 28 JUN 1991
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Greene, Jamie L.
CC REGISTRATION NUMBER: 32,467
CC REFERENCE/DOCKET NUMBER: 07362-0101
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 404-818-3700
CC TELEFAX: 404-818-3799
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 2205 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC HYPOTHETICAL: NO
CC ANTI-SENSE: NO
CC FRAGMENT TYPE: N-terminal
CC ORIGINAL SOURCE:
CC ORGANISM: Rubella virus
CC STRAIN: Thierlen
CC SEQUENCE 2205 AA; 24023 MW; 22500019 CN;
SO
Query Match 3.6%; Score 102; DB 1; Length 2205;
Best Local Similarity 33.3%; Pred. No. 2.44e+01;
Matches 37; Conservative 17; Mismatches 48; Indels 9; Gaps 7;
DB 2086 GARPPPHHRRARDSDP-LRAOSPRRLTPLYVGLTTLRSSPTVSPHLVGTQ 2144
QY 9 GPRSSGGARARARCARCPRAROSPAR-LTP--DT-VLVDLVSDSDEEV-LEVADPV 62
DB 2145 LLPFGARCGCPGFYYPHHG-GPEGPRGTIPRPARGTRRRRLABAPAA 2194
QY 63 EYVAVRLPAKAPEDSDSDSGAAGPAGA-PRTLVRRRRRLADPGCAP 112
RESULT 9
ID US-08-726-306A-56 STANDARD; PRT: 181 AA.
XX xxxxxx
AC
XX
DT
XX
DE
XX
Sequence 56, Application US/08726306A
CC Sequence 56, Application US/08726306A
CC Patent No. 5958684
CC GENERAL INFORMATION:
CC APPLICANT: van Leeuwen, Frederik Willem
CC APPLICANT: Burbach, Johannes Peter Henri

CC APPLICANT: Grosveld, Franklin G.
CC TITLE OF INVENTION: DIAGNOSIS METHOD AND REAGENTS
CC NUMBER OF SEQUENCES: 189
CC CORRESPONDENCE ADDRESSES:
CC ADDRESSEE: Banner & Witcoff, Ltd.
CC STREET: 1 Financial Center
CC CITY: Boston
CC STATE: MA
CC COUNTRY: US
CC ZIP: 02111
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Wordperfect 6.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/726.306A
CC FILING DATE: 02-Oct-1996
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: GB 95/20080.4
CC FILING DATE: 02-Oct-1995
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 60/009,832
CC FILING DATE: 01-Jan-1996
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Williams, Ph.D., Kathleen M.
CC REGISTRATION NUMBER: 34,380
CC REFERENCE/DOCKET NUMBER: 96,048-A (3255/00784)
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 345-9100
CC TELEFAX: (617) 345-9111
CC INFORMATION FOR SEQ ID NO: 56:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 181 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: unknown
CC MOLECULE TYPE: peptide
CC SEQUENCE 181 AA; 17613 MW; 98529 CN;
SO
Query Match 3.5%; Score 100; DB 2; Length 181;
Best Local Similarity 55.2%; Pred. No. 3.21e+01;
Matches 16; Conservative 6; Mismatches 5; Indels 2; Gaps 2;
DB 42 PLRGRAYRARGPRDARG-AAPGRARS 69
QY 4 PLR-GRGRSRNGRARGARARGRCRPA 31
RESULT 10
ID PCT-US96-10602-6 STANDARD; PRT: 397 AA.
XX xxxxxx
AC
XX
DT
XX
DE
XX
Sequence 6, Application PC/TUS9610602
CC Sequence 6, Application PC/TUS9610602
CC GENERAL INFORMATION:
CC APPLICANT: The General Hospital Corporation
CC TITLE OF INVENTION: INHIBITION OF HEPATITIS B REPLICATION
CC NUMBER OF SEQUENCES: 14
CC CORRESPONDENCE ADDRESSES:
CC ADDRESSEE: Fish & Richardson P.C.
CC STREET: 225 Franklin Street
CC CITY: Boston
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02110-2804
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/10602
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/017,814
FILING DATE: 20-JUN-1995
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00786/282001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 397 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE 397 AA; 45191 MW; 946620 CN;

Query Match 3.5%; Score 99; DB 3; Length 397;
Best Local Similarity 48.7%; Pred No. 3,69e+01;
Matches 19; Conservative 7; Mismatches 9; Indels 4; Gaps 3;

Db 152 RGRSPRRRT-PSPPR-R--RSQSPRRRRSSQSLGPLEVL 186
|||:||| :||| |::|||:-||| ||
QY 6 RGRGPRSRGRCARRARAGRCPRARQSAPRLIPDTVL 44

RESULT 11
ID PCT-US94-04496-30 STANDARD: PRT; 568 AA.
AC xxxxxx
XX
XT
DE Sequence 30, Application PC/TUS9404496
XX
XX GENERAL INFORMATION:
CC APPLICANT: Croce, Carlo
CC TITLE OF INVENTION: Diagnostics, Therapeutics and Methods
CC TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias
CC TITLE OF INVENTION: Resulting from Chromosome Abnormalities In the ALL-1
CC NUMBER OF SEQUENCES: 86
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Woodcock, Washburn, Kurtz, Macklewitz &
CC ADDRESSEE: Norris
CC STREET: One Liberty Place, 46th floor
CC City: Philadelphia
CC STATE: Pennsylvania
CC COUNTRY: USA
CC ZIP: 19103
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: floppy disk
CC OPERATING SYSTEM: IBM PC compatible
CC SOFTWARE: Patentin Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US94/04496
CC FILING DATE:
CC CLASSIFICATION:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Deluca Esq., Mark
CC REGISTRATION NUMBER: 33,229
CC REFERENCE/DOCKET NUMBER: TJU-1242

[illegible]

```
CC      REGISTRATION NUMBER: 33,229
CC      REFERENCE/DOCKET NUMBER: TJU-0855
CC      TELECOMMUNICATION INFORMATION:
CC      TELEPHONE: (215) 568-3100
CC      TELEFAX: (215) 568-3439
CC      INFORMATION FOR SEQ ID NO: 30:
CC      SEQUENCE CHARACTERISTICS:
CC          LENGTH: 568 amino acids
CC              TYPE: amino acid
CC              TOPOLOGY: linear
CC      MOLECULE TYPE: protein
CC      SEQUENCE 568 AA; 63367 MW; 1762870 CN;
SO
Dd      Query Match 3.5%; Score 99; DB 1; Length 568;
        Best Local Similarity 26.7%; Pred. No. 3,69e+01;
        Matches 32; Conservative 32; Mismatches 49; Indels 7; Gaps 7;
Oy      132 RRKLKAGDDPRSTHTSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSST 191
        ||||| : | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy      102 RRRLLDPGEAPVYPPYSGVQSSLTLPDNSSLKLCPEPEDEADLTNGSSSPSDDL 161
        : : : : : | : | : | : : : : : : : : : : : : : : : : : : : : : : : : :
Oy      192 SFSKP-HK-LMKE-HKE-KPSKDSREHK-SAFKEPS-RDHNKSSEKSKPKK-NRPLKE 244
        : : : : : | : | : | : : : : : : : : : : : : : : : : : : : : : : : : :
Oy      162 PGSGWRRKKLRKKCKEKKMEFPDQDISPLPQPSSRKRSKRHTALQKLREVNNRLD 221
        : : : : : | : | : | : : : : : : : : : : : : : : : : : : : : : : : : :
Oy
RESULT 13 STANDARD; PRT; 30 AA.
ID US-08-218-608-10
xx xxxxxx
xx
xx
xx
xx
xx
xx
xx
Sequence 10, Application US/08218608
DE
cc Sequence 10, Application us/08218608
cc Patent No. 5607859
cc GENERAL INFORMATION:
cc APPLICANT: BIEMANN, KLAUS
cc APPLICANT: JUHASZ, PETER
cc TITLE OF INVENTION: METHODS AND PRODUCTS FOR MASS
cc TITLE OF INVENTION: SPECTROMETRIC MOLECULAR WEIGHT DETERMINATION OF POLYIONIC
cc TITLE OF INVENTION: ANALITES EMPLOYING POLYIOTNIC REAGENTS
cc NUMBER OF SEQUENCES: 12
cc CORRESPONDENCE ADDRESS:
cc ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
cc STREET: 600 ATLANTIC AVENUE
cc CITY: BOSTON
cc STATE: MA
cc COUNTRY: USA
cc ZIP: 02210
cc COMPUTER READABLE FORM:
cc MEDIUM TYPE: Floppy disk
cc COMPUTER: IBM PC compatible
cc OPERATING SYSTEM: PC-DOS/MS-DOS
cc SOFTWARE: Patentin Release #1.0, Version #1.25
cc CURRENT APPLICATION DATA:
cc APPLICATION NUMBER: US/08/218,608
cc FILING DATE: 28-MAR-1994
cc CLASSIFICATION: 436
cc ATTORNEY/AGENT INFORMATION:
cc NAME: GATES, EDWARD R.
cc REGISTRATION NUMBER: 31,616
cc REFERENCE/DOCKET NUMBER: M0656/7013
cc TELECOMMUNICATION INFORMATION:
cc TELEPHONE: 617-720-3500
cc TELEFAX: 617-720-2441
cc INFORMATION FOR SEQ ID NO: 10:
cc SEQUENCE CHARACTERISTICS:
cc LENGTH: 30 amino acids
cc type: amino acid
cc STRANDEDNESS: single
cc TOPOLOGY: linear
```

```

CC MOLECULE TYPE: Peptide
CC HYPOTHETICAL: NO
CC ANTI-SENSE: NO
CC ORIGINAL SOURCE:
CC ORGANISM: NONE (SYNTHETIC PEPTIDE)
SQ SEQUENCE 30 AA; 3217 MW; 2370 CN;

Query Match 3.4%; Score 98; DB 1; Length 30;
Best Local Similarity 62.1%; Pred.No. 4.23e+01;
Matches 18; Conservative 4; Mismatches 3; Indels 4; Gaps 4;

Db 1 RGRG-RGKRGGRGRGRGRGRGR-RGR 27
||| |:|| ||| |:|| ||| |:||
QY 6 RGRGRSRNG-RGRARRARG-ARGRCPRAR 32

RESULT 14 STANDARD; PRT; 123 AA.
ID US-08-306-078-5
XX xxxxxx
D1
D7
XX
DE Sequence 5, Application US/08306078
XX
CC Sequence 5, Application US/08306078
CC Patent No. 5827646
CC GENERAL INFORMATION:
CC APPLICANT: Middeldorp, Jaap Michiel JM
CC APPLICANT: van Grunsven, Mouterus Marinus Johannes WMJ
CC TITLE OF INVENTION: Diagnostic reagents for the
CC TITLE OF INVENTION: detection of antibodies to EBV.
CC NUMBER OF SEQUENCES: 9
CC CORRESPONDENCE ADDRESSES:
CC ADDRESSSEE: AKZO NOBEL PHARMA
CC STREET: 1330 Piccard Drive
CC CITY: Rockville
CC STATE: Maryland
CC COUNTRY: USA
CC ZIP: 20850-4377
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/306,078
CC FILING DATE: 14-SEP-1994
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: EP 93202659.4
CC FILING DATE: 14-SEP-1993
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Blackstone, William B.
CC REGISTRATION NUMBER: 29,772
CC INFORMATION FOR SEQ ID NO: 5:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 123 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 123 AA; 12990 MW; 70971 CN;

Query Match 3.4%; Score 97; DB 2; Length 123;
Best Local Similarity 28.3%; Pred.No. 4.85e+01;
Matches 28; Conservative 26; Mismatches 38; Indels 7; Gaps 7;

Db 21 RARG-RGRG-GGERPRSPSSSSSSPPRRPGRPFHFYGEADYEYHOGGPDG 78
|::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
QY 6 RGRGRSRNGRARRARGRCPRARNSPARLIPDT-VLVLDVSDSE-EYLEVDAPVE 63

. 79 EP-D-VPPGAIEGGPADDPGEGETGTGPRGGDG-GRRKK 114

```

OY 64 VPAVRLPAPAKPEODSDSDSEGAAGPAGAPRTLVRRRR 102

RESULT 15
ID US-08-240-717A-1 STANDARD: PRT; 123 AA.
XX
AC
XX
DT
XX
DE
XX

Sequence 1, Application US/08240717A

CC Sequence 1, Application US/08240717A
CC Patent No. 5965353

CC GENERAL INFORMATION:
CC APPLICANT: Middelorp, Jaap M.
CC TITLE OF INVENTION: Epstein Barr Virus Peptides and
CC TITLE OF INVENTION: Antibodies Against These Peptides
CC NUMBER OF SEQUENCES: 6

CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Akzo No. 5965353e1 Patent Dept.

CC STREET: 1300 Piccard Drive, Suite 206
CC CITY: Rockville
CC STATE: Maryland
CC COUNTRY: US

CC ZIP: 20850

CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Floppy disk

CC COMPUTER: IBM PC compatible

CC OPERATING SYSTEM: PC-DOS/MS-DOS

CC SOFTWARE: Patentin Release #1.0, Version #1.30

CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: US/08/240,717A

CC FILING DATE: 11-MAY-1994

CC CLASSIFICATION: 435

CC PRIOR APPLICATION DATA:

CC APPLICATION NUMBER: EP 92.202797.4

CC FILING DATE: 14-SEP-1992

CC ATTORNEY/AGENT INFORMATION:

CC NAME: Gormley, Mary E.

CC REGISTRATION NUMBER: 34,409

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: (301) 258-5200

CC TELEFAX: (301) 977-0847

CC INFORMATION FOR SEQ ID NO: 1:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 123 amino acids

CC TYPE: amino acid

CC STRANDEDNESS: single

CC TOPOLOGY: linear

CC MOLECULE TYPE: peptide

CC SEQUENCE 123 AA; 12990 MM; 70971 CN;

Query Match 3.4%; Score 97; DB 2; Length 123;

Best Local Similarity 28.3%; Pred. No. 4.85e+01;

Matches 28; Conservative 26; Mismatches 38; Indels 7; Gaps 7;

DB 21 RARG-RGRG-RGRPRRSPSSSSGSPRRPPGRPFPHYGEADYFEYHOEGPDPG 78
QY 6 RGRPRRSGRGARARARARCRPAROSPARLPDT-VLVLDVSDSE-EVLEVADPVE 63
DB 79 EP-D-VPPGAIEGGPADDPGEGPSTGPRGQDG-GRRRK 114
QY 64 VPAVRLPAPAKPEODSDSDSEGAAGPAGAPRTLVRRRR 102

Search completed: Fri Feb 18 12:02:29 2000
Job time : 27 secs.

(TM)

Distribution rights by Oxford Molecular Ltd

ot generated

1 МАЕРІ

Gap 11

122810 seqs, 40068593 residues

Listing first 45 summaries

1:plr1 2:plr2 3:plr3 4:plr4

Mean 49.563; Variance 122.566; scale 0.404

and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	% Match	Query Length	DB	ID	Description	Pred. No.
	1	161	5.6	95	2	JC4760	SM3 protein - human	2.27e-06
	2	152	5.3	91	2	JC5582	SM3 protein - Caenor	2.96e-05
	3	129	4.5	1402	2	I46707	initiation factor 4-g	1.54e-0204
	4	125	4.4	101	2	S63999	SM3 protein - yeast	4.34e-02
	5	117	4.1	771	2	B70564	hypothetical protein	3.25e-01
	6	113	4.0	673	2	S60173	fragile X mental reta	8.63e-01
	7	114	4.0	932	2	A42632	cell adhesion molecucl	6.77e-01
	8	115	4.0	1870	2	S37671	bat2 protein - human	5.31e-01
	9	113	4.0	1872	2	S36152	bat2 protein - human	5.31e-01
	10	113	4.0	2142	2	B35098	MHC class. III histoc	8.63e-01
	11	111	3.9	359	1	I51734	transcription factor	1.40e+00
	12	111	3.9	1030	2	S11034	gene p1 protein - fru	1.40e+00
	13	108	3.8	229	1	WMA051	late 3K protein - hu	2.84e+00
	14	109	3.8	340	1	B55973	transcription factor	2.24e+00
	15	109	3.8	380	2	T00792	hypothetical protein	2.24e+00
	16	110	3.8	385	2	I38498	GTP cyclohydrolase II	1.77e+00
	17	109	3.8	431	2	E70579	probable mure protein	2.24e+00
	18	109	3.8	430	2	S09824	hypothetical protein	2.24e+00
	19	110	3.8	476	2	S36513	I2 protein - human pa	1.77e+00
	20	109	3.8	596	2	A45195	adenylylcyase type	2.24e+00
	21	110	3.8	657	2	S05517	lamin - chicken	1.77e+00
	22	109	3.8	1184	2	A42904	adenylylcyase type	2.24e+00
	23	108	3.8	1210	2	A48001	phospholipase C (EC 3	2.84e+00

45	104	3.6	265	1	LEPRA1B	apolipoprotein A-I pr	7.16e+00
44	106	3.7	3414	1	GNM5WE	eye development prote	4.52e+00
43	105	3.7	1893	2	GNM5WE	eye development prote	4.52e+00
42	106	3.7	1703	2	S133441	EF protein - streptoc	5.69e+00
41	105	3.7	1822	2	S133441	EF protein - yeast	4.52e+00
40	105	3.7	1234	2	I38894	phospholipase C beta-	5.69e+00
39	107	3.7	1107	2	E65226	phospholipase C (EC	3.58e+00
38	105	3.7	1051	2	S27002	phospholipase C (EC	5.69e+00
37	107	3.7	913	2	S20590	exo-alpha-1,4-lidase	4.52e+00
36	106	3.7	676	1	EDB823	immediate-early prote	3.58e+00
35	107	3.7	525	1	EDB822	immediate-early prote	3.58e+00
34	107	3.7	676	1	S36405	nucleocapsid protein	3.58e+00
33	105	3.7	358	1	A55973	transcription factor	5.69e+00
32	106	3.7	349	1	I50369	transcription factor	4.52e+00
31	105	3.7	349	1	I67417	transcription factor	4.52e+00
30	106	3.7	349	1	I53277	transcription factor	4.52e+00
29	106	3.7	349	1	I67418	transcription factor	4.52e+00
28	105	3.7	338	2	I38587	retrovirus-related le	5.69e+00
27	110	3.8	2424	2	I46479	calcium channel BI-2	1.77e+00
26	110	3.8	2424	2	I46480	calcium channel BI-2	1.77e+00
25	110	3.8	2284	1	GNM5V	genome polypotein -	2.84e+00
24	110	3.8	1302	2	A41249	multidrug resistance	1.77e+00
23	118	3.8	1302	2	A41249	multidrug resistance	1.77e+00

ALIGNMENTS

```

ENTRY      1
TITLE      JCA4760
ORGANISM   #type complete
            SMT3 protein - human
            #formal_name Homo sapiens #common_name man
            DATE 10-May-1996 #sequence_revision 16-Aug-1996 #text_change
            17-Mar-1999

ACCESSIONS
REFERENCE   JCA4760
            Mannen, H.; Tseng, H.M.; Cho, C.; Li, S.S.L.
            Blochem. Biophys. Res. Commun. (1996) 222:178-180
            Cloning and expression of human homolog HSMT3 to yeast SMT3
            suppressor of HIF2 mutations in a centromere protein gene
            #accession JCA4760
            #molecule_type mRNA
            #residues 1-95 #label MAN
            #cros-references GB:L76416

GENETICS
            #gene GDB:SMT3H2; HSMT3; SMT3B
            #cros-references GDB:S75262
            #map_position 4
            #classification #superfamily yeast SMT3 protein
            SUMMARY #length 95 #molecular_weight 10671 #checksum 3774

Query Match 5.6%; Score 161; DB 2; Length 95;
Best Local Similarity 32.9%; Pred. No. 2,27e-06;
Matches 25; Conservative 19; Mismatches 31; Indels 1; Gaps 1.

Db 12 TENDHINKVAGDGS-VVOFKRRHTPSLKKAYCEROGSMROIREFPDGPINET 70
OY 335 TETSOELRLKRVQKXKQMEFLISLPSPLKVLMSHIEAMGLSGHKLSEFFDQTKLSGK 394
            ||::: ||:|::: || || | |:: ||::
            :||:|:| |::|
            OY 395 ELPADLIGESGLIEV 410

Db 71 DTPAQLEMEDDITDV 86
OY 395 ELPADLIGESGLIEV 410
            :||:|:| |::|

RESULT 2
ENTRY    JCS582
TITLE    SMT3 protein - Caenorhabditis elegans
ORGANISM #formal_name Caenorhabditis elegans
DATE     23-Sep-1997 #sequence_revision 23-Sep-1997 #text_change
            17-Mar-1999

ACCESSIONS
REFERENCE   JCS582
            Choudhury, B.K.; Li, S.S.L.
            Blochem. Biophys. Res. Commun. (1997) 234:788-792
            #journal #journal

```

##title	Identification and characterization of the SMT3 cDNA and gene from nematode <i>Caenorhabditis elegans</i> .
##cross-references	NCBI:18858
#accession	J05582
##molecule-type	DNA
##residues	1-91 ##label CHO
##cross-references	GB:U94830; NID:g2241094; PID:g2241095
COMMENT	This protein is a suppressor of MIF2 mutation in a centromere protein.
GENETICS	
#gene	smt3
#introns	22/3: 56/2
CLASSIFICATION	#superfamily yeast SMT3 protein
SUMMARY	#length 91 #molecular-weight 10222 #checksum 2101
Query Match	5.3%; Score 152; DB 2; Length 91;
Best Local Similarity	22.5%; Pred. No. 2,95e-05;
Matches	18; Conservative 31; Mismatches 30; Indels 1; Gaps 1;
Db	5 AAOAGNAEYIKKVKYGQSDSEV-HRRVKKGTSMARKKSYADRTGYAVNSLRFLEGGRR 63
Qy	331 SESETTSOELRRVGGKHKMLKLEISLSPDSPLKVLMSHYEFAMGLSGHRSFFDGTR 390
Db	64 INDDPTKLTLEMDPDVIEV 83
Qy	391 LSGKELPADLGLESGDILEY 410
RESULT	3
ENTRY	146707 #type complete
TITLE	Initiation factor 4-gamma - rabbit
ORGANISM	#normal_name Oryctolagus cuniculus #common_name domestic rabbit
DATE	14-Feb-1997 #sequence-revision 14-Feb-1997 #text-change 14-Feb-1997
ACCESSIONS	146707
REFERENCE	146707
#authors	Yan, R.; He, W.; Rhoads, R.E.
#journal	J. Biol. Chem. (1993) 268:19200-19203
#title	Mapping the cleavage site in protein synthesis initiation factor eIF-4 gamma of the 2a proteases from human coxsackievirus and rhinovirus.
#cross-references	NCBI:93374895
#accession	146707
#status	preliminary; translated from GB/EMBL/DBJ
##molecule-type	MRNA
##residues	1-1402 ##label YAN
##cross-references	GB:U2090; NID:g404774; PID:g404775
SUMMARY	#length 1402 #molecular-weight 154049 #checksum 1846
Query Match	4.5%; Score 129; DB 2; Length 1402;
Best Local Similarity	20.8%; Pred. No. 1.54e-02;
Matches	46; Conservative 62; Mismatches 103; Indels 10; Gaps 10;
Db	482 ALESSRGP-PRGGCGGELPRGAAGLGRPRSLQPPRPGCARLLIASYTIEDIKLNAKAKAW 540
Qy	4 PIRGRPRRGCGGARRARGRCPRARQSPARLLPDTLVLDVSDSEVLEVPDVE 63
Db	541 KPSKRTADK-DR-GEEPDGSKTODLFRVRSILINKLTPOMPOQLMQTOTLAIDTEG 598
Qy	64 VVYARLPAPAKPRQDDSDSDEGA-AEGPACAPFTLVARRRRRLIDPEAAVVPY-YSGKY 121
Db	599 ASKGSUTSLRRPQNTQWPSQHVLPFGASATTEKEFTVYN-FRKLINRCQKEFE 657
Qy	122 QGSLLNLIPNNSLLKLCPSF-PEDEADLTNNGSSPDALPGSSPWKKLRKKCKE-E 179
Db	658 KKKDDDEVFEKKOKEMDEAATAEERELKKELEFARIAR 698
Qy	180 K-KME-EFPDODISPLPQSSRNKSRKHTALQKLRVNR 218
RESULT	4
ENTRY	S63999 #type complete

```

TITLE          SMT3 protein - yeast (Saccharomyces cerevisiae)
ALTERNATE_NAMES    protein YDR510W
ORGANISM         #format_name Saccharomycos cerevistae
DATE             16-May-1996 #sequence_revision 19-Jul-1996 #text_change
                04-Sep-1998
ACCESSIONS       563999; 569567
REFERENCE        Melun, P.B.; Koshland, D.E.
#authors         Submitted to the EMBL Data Library, May 1995
#accession       563999
               ##molecule_type DNA
               ##residues      1-101 ##label MEL
               ##cross-references EMBL:U27233; NID:g881371; PID:g881372
               ##experimental_source strain YPH1/YMN214
REFERENCE        S69553
#authors         Dietrich, F.S.
#description     The sequence of S. cerevisiae cosmid 8166, 9787, 9717, and
                 lambda 3073.
#accession       S69567
               ##molecule_type DNA
               ##residues      1-101 ##label DIE
               ##cross-references EMBL:U33057; NID:g927764; PID:g927779; MIPS:YDR510W
GENETICS
#gene            SGD:SMT3
#cross-references SCD:S0002918; MIPS:YDR510W
CLASSIFICATION   #superfamily yeast SMT3 protein
SUMMARY          #length 101 #molecular_weight 11597 #checksum 8485

Query Match           4.4% Score 125: DB 2: Length 101:
Best Local Similarity 33.3% Pred. No.4.34e+02:
Matches              16: Conservative 11: Mismatches 21: Indels 0: Gaps 0:

Db      43 TPLRLMEAFARQKREMSLRFLFYDIRIQADOTPEDLDMDNDIE 90
      :|::||: | ||:|| :: |: | | | | | | |
Oy      362 SPLATMNGHYEAMGSLGRHISFFPDTRKSCKELPADLGLESGDLIE 409


RESULT      5
ENTRY TITLE B70564 #type complete
             hypothetical protein RV3649 - Mycobacterium tuberculosis
             (strain H37RV)
ORGANISM    #format_name Mycobacterium tuberculosis
DATE        17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change
            17-Jul-1998
ACCESSIONS  B70564
REFERENCE    A70500
#authors     Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barrett, I.I., C.E.; Tekle, E.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornby, T.; Jagels, K.; Krogh, A.; McLaren, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skellon, S.; Squares, S.; Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
#journal     Nature (1998) 393:537-544
#title       Deciphering the biology of Mycobacterium tuberculosis from
             the complete genome sequence.
#cross-references MUID:98293987
#accession   B70564
#status      preliminary; nucleic acid sequence not shown; translation not shown

##molecule_type DNA
##residues      1-771 ##label COL
##cross-references GB:I29543; GB:AI123456; NID:g3261770; PID:e316541;
             PID:g2105045
##experimental_source strain H37RV

GENETICS
RGV3649
#gene         length 771 #molecular-weight 81409 #checksum 4348
SUMMARY

```


#journal Nature Genet. (1993) 3:137-145
#title Dense Alu clustering and a potential new member of the NRKappab family within a 90 kilobase HLA class III segment.
#accession S36152
#status preliminary
#molecule_type DNA
#residues 1-1872 ##label IRI
#cross-references EMBL:215025
#note In the authors' translation residues 32-34 are shown after residue 4 and, consequently, residues 5-31 are displaced three codons to the right
#note the authors translated the codon AAT for residue 1000 as His

GENETICS
#introns 38/2; 97/2; 129/3; 154/1; 202/1; 252/3; 279/2; 327/1; 357/2; 429/3; 588/1; 651/1; 751/1; 821/2; 1436/2; 1525/1; 1568/3; 1612/1; 1640/3; 1715/3; 1762/3; 1786/3; 1842/3
#length 1872 #molecular-weight 199338 #checksum 7324

QUERY
#query Match 4.0%; Score 115; DB 2; Length 1872;
Best Local Similarity 29.5%; Pred. No. 5.31e-01;
Matches 31; Conservative 31; Mismatches 37; Indels 6; Gaps 6;

Db 1024 RGRGTFARG-RGRGTFGGRGQANSAYTESFEEMNGVE-VGQGDQTTLLPEAA-L 1080
OY 6 RGRGP-RSRGGGARARAGRCPRAROSPRLIPDVLVDVSDSEVLEVDPEV 64

Db 1081 P-ARHGAVOSMRKSPGAGCAQKQAKAPRMVILQTRRLPH 1124
OY 65 PVARLPAPAKPEQSDSDSEGAAGPACAPRLVRR-RRRLDDP 108

RESULT 10
ENTRY B35098 #type complete
TITLE MHC class IIr histocompatibility antigen HLA-B-associated transcript 2 - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 10-Aug-1990 #sequence_revision 06-Nov-1992 #text_change 10-Sep-1997

ACCESSIONS B35098
REFERENCE A35098
#authors Banerji, J.; Sands, J.; Strominger, J.L.; Spies, T.
#journal Proc. Natl. Acad. Sci. U.S.A. (1990) 87:2374-2378
#title A gene pair from the human major histocompatibility complex encodes large proline-rich proteins with multiple repeated motifs and a single ubiquitin-like domain.
#cross-references MUID:90192810
#accession B35098
#status preliminary
#molecule_type mRNA
#residues 1-2142 ##label BAN
#cross-references GB:M33509; NID:g179338; PID:g179339; GB:M31293
#note The authors translated the codon AGT for residue 97 as Gly

SUMMARY
#length 2142 #molecular-weight 227840 #checksum 5704

QUERY Match 4.0%; Score 113; DB 2; Length 2142;
Best Local Similarity 29.5%; Pred. No. 8.63e-01;
Matches 31; Conservative 30; Mismatches 38; Indels 6; Gaps 6;

Db 1012 RGRGTFARG-RGRGTFGGRGQANSAYTESFEEMNGVE-VGQGDQTTLLPEAA-M 1068
OY 6 RGRGP-RSRGGGARARAGRCPRAROSPRLIPDVLVDVSDSEVLEVDPEV 64

Db 1069 P-ARHGAVOSMRKSPGAGCAQKQAKAPRMVILQTRRLPH 1112
OY 65 PVARLPAPAKPEQSDSDSEGAAGPACAPRLVRR-RRRLDDP 108

RESULT 11
ENTRY 151734 #type complete
TITLE transcription factor Isl-2 - zebra fish
ALTERNATE_NAMES insulin enhancer-binding protein Isl-2; Isl-2 protein

ORGANISM #formal_name Brachydanio rerio #common_name zebra fish
DATE 13-Mar-1997 #sequence_revision 13-Mar-1997 #text_change 05-Sep-1997

ACCESSIONS 151734; S57407
REFERENCE 151734
#authors Tokumoto, M.; Gong, Z.; Tsubokawa, T.; Hew, C.L.; Uemura, K.; Hotta, Y.; Okamoto, H.
#journal Dev. Biol. (1995) 171:578-589
#title Molecular heterogeneity among primary motoneurons and within motomes revealed by the differential mRNA expression of novel Isl-1 homologs in embryonic zebrafish.
#cross-references MUID:9605022
#accession 151734
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-359 ##label TOK
#cross-references GB:D38453; NID:g1037165; PID:g1037166

REFERENCE S57407
#authors Appel, B.; Korzh, V.; Glasgow, E.; Thor, S.; Edlund, T.; Dawid, I.; Eisen, J.
#submission submitted to the EMBL Data Library, June 1995
#description Motoneuron fate specification and patterned LIM homeobox gene expression in embryonic zebrafish.
#accession S57407
#status preliminary
#molecule_type mRNA
#residues 1-359 ##label APP
#cross-references EMBL:X88805; NID:g871000; PID:g871001

GENETICS
#gene Isl-2
CLASSIFICATION #superfamily transcription factor Isl-1; homeobox homology; LIM metal-binding repeat homology
KEYWORDS DNA binding; duplication; homeobox; nucleus; transcription regulation; zinc

FEATURE
27-80 #domain LIM metal-binding repeat homology #label LIM1
89-142 #domain LIM metal-binding repeat homology #label LIM2
192-248 #domain homeobox homology #label HOX

SUMMARY
#length 359 #molecular-weight 40227 #checksum 7994

QUERY Match 3.9%; Score 111; DB 1; Length 359;
Best Local Similarity 30.3%; Pred. No. 1.40e+00;
Matches 23; Conservative 23; Mismatches 24; Indels 6; Gaps 6;

Db 174 PVROPFRHHVHKSEKTRVRYLVNKEQLTQTCYANRPD-ALM-KE-QLVEMTG- 229
OY 192 PLPQSSNRKSRKHEALQKREV-N-KRLQDLRSLSPKHQSPALQSTDEVVLVGP 249

Db 230 LSPRYRYVFNKKRCK 245
OY 250 VLPOSSRLFTKIRCR 265

RESULT 12
ENTRY S11034 #type complete
TITLE gene P1 protein - fruit fly (Drosophila melanogaster)
ORGANISM #formal_name Drosophila melanogaster
DATE 21-Nov-1993 #sequence_revision 01-Dec-1995 #text_change 24-Sep-1998

ACCESSIONS S11034
REFERENCE S11034
#authors Maschut, F.; Dubertret, M.L.; Therond, P.; Claverie, J.M.; Lepesant, J.A.
#journal J. Mol. Biol. (1990) 214:359-372
#title Structure of the ecdysone-inducible P1 gene of Drosophila melanogaster.
#cross-references MUID:90339481
#accession S11034
#status preliminary
#molecule_type DNA
#residues 1-1030 ##label MAS
#cross-references EMBL:X69965; NID:g7960; PID:g7961

GENETICS

#gene FlyBase:FBp1
#cross-references FlyBase:FBgn0000639
#introns 31/1
SUMMARY length 1030 #molecular-weight 119364 #checksum 1902

Query Match 3.9%; Score 111; DB 2; Length 1030;
Best Local Similarity 20.9%; Pred. No. 1.40e+00;
Matches 42; Conservative 66; Mismatches 76; Indels 17; Gaps 15;

Db 559 TDDLELHMQRRQRLQKH-QNDDDD-DD-NDDVNVYHROGLSRSLPMLRQONNL 615
OY 156 SEDDLPLPS-GSPWRRKRLKRCCKEKKMEFPDIDISLPQSSSRKSKRTALQKLE 214
Db 616 SEIVAHNRQ-LVARINQESIAGOLIEEOQLINPRLQSEYALRLNQTIRINSORR 674
OY 215 VKRRQDLRSCSLPKHQSPALQS--TDDEVYLVESPVLPQSSRL-FTLK-TRCRADLVR 270
Db 675 OVLAIQIGIQRIQVIGVLSQVNVNSLRGVIDQROVESLADVLGRLQGVGMTII 734
OY 271 --LP-V-RMSEPLQNVVDHMANHLGVSPNRIILFG-ESE-LSPATPSTL-KLGADII 323
OY 735 ROYVDNNSQIDRRNGLGIRL 755
OY 324 DCVVLASSEATE-TSQELRL 343

RESULT 13
ENTRY WMA051 #type complete
TITLE late 33k protein - human adenovirus 5
ORGANISM #formal_name Mastadenovirus h5 #common_name human adenovirus 5
#note host Homo sapiens (man)
DATE 30-Sep-1992 #sequence-revision 30-Sep-1992 #text-change 04-Mar-1994

ACCESSIONS F39449
REFERENCE A39449
#authors Chroboczek, J.; Bieber, F.; Jacrot, B.
#journal Virology (1992) 186:280-285
#title The sequence of the genome of adenovirus type 5 and its comparison with the genome of adenovirus type 2.
#cross-references MUID:92087470
#accession F39449
#molecule_type DNA
#residues 1-229 #label CHR
#cross-references GB:M73260
CLASSIFICATION #superfamily adenovirus late 33k protein
KEYWORDS late protein
SUMMARY #length 229 #molecular-weight 25164 #checksum 9588

Query Match 3.8%; Score 108; DB 1; Length 229;
Best Local Similarity 31.6%; Pred. No. 2.84e+00;
Matches 18; Conservative 14; Mismatches 24; Indels 1; Gaps 1;

Db 42 IDEASEVEEVSDETPSPVAFSPA-POKSATGSMATTSAPAPALVRRPNRR 97
OY 48 VSDSEVLEVADPEVYARLPAPAKPEQSDSDSEGAABEPAGAPRLVRRRRR 104

RESULT 14
ENTRY B55973 #type fragment
TITLE transcription factor isl-2b - chinook salmon (fragment)
ALTERNATE_NAMES insulin enhancer-binding protein isl-2b; isl-2b protein
ORGANISM #formal_name Oncorhynchus tshawytscha #common_name chinook salmon
DATE 03-Oct-1995 #sequence-revision 03-Oct-1995 #text-change 05-Sep-1997
ACCESSIONS B55973; S52091; S19956
REFERENCE A55973
#authors Gong, Z.; Hul, C.; Hew, C.L.
#journal J. Biol. Chem. (1995) 270:3335-3345
#title Presence of Isl-1-related LIM domain homeobox genes in teleost and their similar patterns of expression in brain and spinal cord.

#accession B55973
#molecule_type mRNA
#residues 1-340 #label GON
#cross-references EMBL:X64884; NID:g64208; PID:g64209
#note authors translated the codon CTG for residue 251 as Met, and ATG for residue 256 as Leu

REFERENCE S52089
#authors Gong, Z.; Hew, C.L.
#journal Biochim. Biophys. Acta (1995) 1260:349-354
#title Several splicing variants of isl-1 like genes in the chinook salmon (Oncorhynchus tshawytscha) encode truncated transcription factors containing a complete LIM domain.
#cross-references MUID:95178560
#accession S52091
#status preliminary
#molecule_type mRNA
#residues 1-340 #label G02
#cross-references EMBL:X64884; NID:g64208; PID:g64209
#note the authors translated the codon CAG for residue 105 as His, GTG for residue 107 as Leu, CTG for residue 251 as Met, and ATG for residue 256 as Leu

GENETICS
#gene isl-2b
CLASSIFICATION #superfamily transcription factor isl-1; homeobox homology; LIM metal-binding repeat homology
KEYWORDS DNA binding; duplication; homeobox; nucleus; transcription regulation; zinc

FEATURE 9-62
71-124 #domain LIM metal-binding repeat homology #label LIM2
173-229 #domain homeobox homology #label HOX
SUMMARY #length 340 #checksum 5243

Query Match 3.8%; Score 109; DB 1; Length 340;
Best Local Similarity 28.9%; Pred. No. 2.24e+00;
Matches 24; Conservative 24; Mismatches 29; Indels 6; Gaps 6;

Db 148 FHIADEVSVRPPHNRHVAKROSEKTRVTVLNEQLHTLRFQCYANRPD-ALM-KE-Q 204
OY 185 FPDQDISPLPQSSSRKSKRTALQKREV-N-KRLQDLRSCSLPKHQSPALQSTDD 242
Db 205 LVEMTG-LSPRVIRVWFQNRCK 226
OY 243 VLVGEGPVLPQSSRLFTLKIRCR 265

RESULT 15
ENTRY T00792 #type complete
TITLE hypothetical protein F24L7.10 - Arabidopsis thaliana
ORGANISM #formal_name Arabidopsis thaliana #common_name mouse-ear cress
DATE 12-Feb-1999 #sequence-revision 12-Feb-1999 #text-change 12-Feb-1999

ACCESSIONS T00792
REFERENCE Z14204
#authors Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul, S.; Mason, T.M.; Kerlavage, A.R.; Adams, M.D.; Somerville, C.R.; Venter, J.C.
#submission submitted to the EMBL Data Library, February 1998
#description Arabidopsis thaliana chromosome II BAC F24L7 genomic sequence.
#accession T00792
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type DNA
#residues 1-380 #label R0U
#cross-references EMBL:AC003974; NID:g2914688; PID:g2914697
#experimental_source cultivar Columbia

GENETICS
#map_position 2
#introns 55/3; 116/1; 161/3; 237/3; 287/3; 357/3
#note F24L7.10
SUMMARY #length 380 #molecular-weight 42844 #checksum 4620

 WISE RELEASE

 (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

Msrch_pp protein - protein database search, using Smith-Waterman algorithm
 Run on: Fri Feb 18 11:58:36 2000; MasPar time 13.08 Seconds
 890.539 Million cell updates/sec
 Modular output not generated.

Title: >US-08-755-584-2
 Description: (1-412) from 5858711.pep
 Perfect Score: 2858
 Sequence: 1 MAEPLRGGRPRSRGNGARR.....GKELPADLGESDLELVWG 412

Scoring table:
 PAM 150
 Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: swiss-prot37
 1:swissprot

Statistics: Mean 50.879; Variance 107.959; scale 0.471

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	161	5.6	95	1	SM32_HUMAN	8.16e-08
2	159	5.6	100	1	SMT3_ORYSA	1.58e-07
3	160	5.6	104	1	SMT3_ARATH	1.14e-07
4	152	5.3	91	1	SMT3_CAEEL	1.55e-06
5	149	5.2	101	1	SMT3_HUMAN	4.07e-06
6	141	4.9	103	1	SM31_HUMAN	5.10e-05
7	129	4.5	1402	1	IF4G_RABIT	1.98e-03
8	126	4.4	90	1	SMT3_SCHPO	4.80e-03
9	125	4.4	101	1	SMT3_YEAST	6.43e-03
10	113	4.0	673	1	FXR2_HUMAN	1.93e-01
11	113	4.0	2142	1	BAT2_HUMAN	1.93e-01
12	111	3.9	359	1	ISL2_BRARE	3.33e-01
13	112	3.9	454	1	VE2_HPV37	2.33e-01
14	112	3.9	707	1	RHO_STRLI	2.53e-01
15	111	3.9	1030	1	FBP1_DROME	3.33e-01
16	108	3.8	229	1	V33P_ADE05	7.43e-01
17	110	3.8	267	1	RS2_DROME	4.35e-01
18	109	3.8	385	1	IS2B_ONCTS	5.69e-01
19	110	3.8	385	1	GCH2_AOBR	4.35e-01
20	109	3.8	431	1	UL61_HCMVA	5.69e-01
21	110	3.8	476	1	VL2_HPV32	4.35e-01
22	109	3.8	619	1	BCHD_CHLVI	5.69e-01
23	110	3.8	657	1	LAMA_CHICK	4.35e-01

RESULT ID	SM32_HUMAN	STANDARD:	PRT:	95 AA.	ALIGNMENTS
AC	P55855;				
DT	01-NOV-1997 (REL. 35, CREATED)				
DT	01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)				
DT	15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)				
DE	UBIQUITIN-LIKE PROTEIN SMT3B (SENTRIN 2).				
GN	SMT3B2 OR SMT3B.				
OS	HOMO SAPIENS (HUMAN), AND BOS TAURUS (BOVINE).				
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;				
OC	PRIMATES; CATARRHINI; HOMINIDAE; HOMO.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	SPECIES=HUMAN; TISSUE=BRAIN;				
RX	MEDLINE: 97237059.				
RA	LAPENTA V., CHIURAZZI P., VAN DER SPEK P.J., PIZZUTI A.,				
RA	HANOKA F., BRAHE C.;				
RT	"SMT3A", a human homologue of the S. cerevisiae Smt3 gene, maps to				
RT	chromosome 21qter and defines a novel gene family.";				
RL	GENOMICS 40:362-367(1997).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	SPECIES=BOVINE; TISSUE=ENDOMETRIUM;				
RA	OF F., BERRY E.S.;				
RL	SUBMITTED (MAR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.				
CC	-1- SIMILARITY: TO UBIQUITIN. BELONGS TO THE SMT3 SUBFAMILY.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sdb.ch/announce/				
CC	or send an email to license@sdb-sdb.ch).				
CC	-----				
DR	EMBL: X95985; E274602; -				
DR	EMBL: U89439; G1888536; -				
DR	MIM: 603042; -				
DR	PROSITE: P550053; UBIQUITIN_2; 1.				
FT	DOMAIN 16 95 UBIQUITIN-LIKE.				
SO	SEQUENCE 95 AA: 10871 MW: E49P30B9 CRC32:				

Query Match 5.6%; Score 161; DB 1; Length 95;
 Best Local Similarity 32.9%; Pred. No. 8.16e-08;
 Matches 25; Conservative 19; Mismatches 31; Indels 1; Gaps 1;
 DB 12 TENNHINKVAGGDS-VVQFKIKRHTPLSKIMAYCERGGISMKQIFRFDGQINFT 70

```

QY 335 TETSOELRLRVQGEKHEKQMLEISLSPDSPKLVMSHYEANGLSGKHSFFDGTKLSGK 394
DB 71 DTPADLEMEDDIDY 86
QY 395 ELPADLGLESGLIEV 410

RESULT 2
ID SMT3_ORYSA STANDARD; PRT: 100 AA.
AC P55853;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE UBIOUITIN-LIKE PROTEIN SMT3.
GN SMT3.
OS ORYZA SATIVA (RICE).
OC EUKARYOTA; VIRIDIPHYTES; STREPTOPHYTA; EMAROPHYTA; TRACHEOPHYTA;
EUPHYLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; LILIOPSIDA; POALES;
OC PACEAE; ORYZA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 97237059.
RA LAPENTA V., CHIRAZZI P., VAN DER SPEK P.J., PIZZUTI A.,
RA HANAOKA F., BRAHE C.;
RT "SMT3A, a human homologue of the S. cerevisiae SMT3 gene, maps to
RT chromosome 21qter and defines a novel gene family.";
RL GENOMICS 40:362-367(1997).
CC -1- SIMILARITY: TO UBIOUITIN. BELONGS TO THE SMT3 SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X99608; E259739;
DR PROSITE: PSS0053; UBIOUITIN_2; 1.
FT DOMAIN 19 96 UBIOUITIN-LIKE.
SQ SEQUENCE 100 AA; 10928 MW; 4E9FADF0 CRC32;

Query Match
Best Local Similarity 29.0%; Pred. No. 1.58e-07;
Matches 20; Conservative 23; Mismatches 25; Indels 1; Gaps 1;

21 INLVKGGDGNFV-FRIKRSQTKKIMNAYCDQSYVMNIAIFDGRRLRGEDTPDEL 79
QY 341 LRLVQGEKHEKQMLEISLSPDSPKLVMSHYEANGLSGKHSFFDGTKLSGKELPADL 400
DB 80 EMEDEDEID 88
QY 401 GLESGDLIE 409

RESULT 3
ID SMT3_ARATH STANDARD; PRT: 104 AA.
AC P55853;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE UBIOUITIN-LIKE PROTEIN SMT3.
GN SMT3.
OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
OC EUKARYOTA; VIRIDIPHYTES; STREPTOPHYTA; EMAROPHYTA; TRACHEOPHYTA;
EUPHYLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
OC CAPPALES; BRASSICACEAE; ARABIDOPSIS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 97237059.
RA LAPENTA V., CHIRAZZI P., VAN DER SPEK P.J., PIZZUTI A.,
RA HANAOKA F., BRAHE C.;

```

```

RT "SMT3A, a human homologue of the S. cerevisiae SMT3 gene, maps to
RT chromosome 21qter and defines a novel gene family.";
RL GENOMICS 40:362-367(1997).
CC -1- SIMILARITY: TO UBIOUITIN. BELONGS TO THE SMT3 SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X99609; E259962;
DR PROSITE: PSS0053; UBIOUITIN_2; 1.
FT DOMAIN 16 93 UBIOUITIN-LIKE.
SQ SEQUENCE 104 AA; 11654 MW; CD706439 CRC32;

Query Match
Best Local Similarity 25.7%; Pred. No. 1.14e-07;
Matches 19; Conservative 26; Mismatches 28; Indels 1; Gaps 1;

13 DCGAHINLVKGGDGNFV-FRIKRSQTKKIMNAYCDQSYVMNIAIFDGRRLRGED 71
QY 336 ETSQELRLRVQGEKHEKQMLEISLSPDSPKLVMSHYEANGLSGKHSFFDGTKLSGK 395
DB 72 TPDELDEMEDDIED 85
QY 396 LPADLGLESGLIE 409

RESULT 4
ID SMT3_CAEEL STANDARD; PRT: 91 AA.
AC P55853;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE UBIOUITIN-LIKE PROTEIN SMT3.
GN SMT3.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 97237059.
RA LAPENTA V., CHIRAZZI P., VAN DER SPEK P.J., PIZZUTI A.,
RA HANAOKA F., BRAHE C.;
RT "SMT3A, a human homologue of the S. cerevisiae SMT3 gene, maps to
RT chromosome 21qter and defines a novel gene family.";
RL GENOMICS 40:362-367(1997).
GN SMT3.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 97318858.
RA CHROUDHURY B.K., LI S.S.;
RT "Identification and characterization of the SMT3 cDNA and gene from
RT nematode Caenorhabditis elegans.";
RL BIOCHEM. BIOPHYS. RES. COMMUN. 234:788-791(1997).
CC -1- SIMILARITY: TO UBIOUITIN. BELONGS TO THE SMT3 SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X99600; E259960;
DR PROSITE: PSS0053; UBIOUITIN_2; 1.
FT DOMAIN 13 91 UBIOUITIN-LIKE.
SQ SEQUENCE 91 AA; 10222 MW; F8E538FD CRC32;

```

```

Query Match:          5.3%;   Score 152; DB 1; Length 91;
Batch Local Similarity 22.5%; Pred. 1.55e-06;
Matches      18; Conservative    31; Mismatches     30; Indels       1; Gaps         1.

Db            5 AAGADNAEYIKIVKVGQDSNEV-HFRVKYTSMAKLKRSYADRGAVNSRLFLPDGR 63
Oy           331 SSENATSTGQLRLRNQGKEKHOMLEISLSPSLKYVMASHHEAMGLSGHKLSTFFFDGT 390
               :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db            64 INDDTPTLEMEDDVIIEV 83
Oy           391 LSGKELPADLGLESGLIIEV 410
               :.:|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|

RESULT        5 STANDARD; PRT; 101 AA.
ID            SM33_HUMAN
AC            Q93068; P55856;
DT            01-NOV-1997 (REL. 35, CREATED)
DT            01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT            15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE            UBIQUITIN-LIKE PROTEIN SMT3C (UBIQUITIN-HOMOLOGY DOMAIN PROTEIN PIC1)
DE            (UBIQUITIN-LIKE PROTEIN UBL1) (UBIQUITIN-RELATED PROTEIN SUMO-1)
DE            (GAP MODIFYING PROTEIN 1) (GMP1) (SENTRIN).
OS            SMT3H3 OR SMT3C.
OS            HOMO SAPIENS (HUMAN), AND MUS MUSCULUS (MOUSE).
OC            EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUETHERIA;
NC            PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RC            (1)
CC            SEQUENCE FROM N.A.
RX            SPECIES=HUMAN; TISSUE=BRAIN; MEDLINE: 97237059.
RA            LAEBENTA V., CHIBRAZI P., VAN DER SPEK P.J., PIZZUTI A.,
RA            HANAOKA F., BRAHE C.;
RT            "SMT3A, a human homologue of the S. cerevisiae SMT3 gene, maps to
RT            chromosome 21qter and defines a novel gene family.";
RL            GENOMICS 40:362-367(1997).
[2]
RP            SEQUENCE FROM N.A.
RC            SPECIES=HUMAN; TISSUE=PLACENTA; MEDLINE: 96400311.
RA            BOODY M.N., HOWE K., ETKIN L.D., SOLOMON E., FREEMONT P.S.;
RT            "PIC 1, a novel ubiquitin-like protein which interacts with the pML
RT            component of a multiprotein complex that is disrupted in acute
RT            promyelocytic leukaemia.";
RL            ONCOGENE 13:971-982(1996).
[3]
RN            SEQUENCE FROM N.A.
RP            SPECIES=HUMAN; MEDLINE: 96411684.
RX            SHEN Z., PARDINGTON-PURTYMUN P.E., COMEAUX J.C., MOYZIS R.K.,
RN            CHEN D.J.;
RT            "UBL1, a human ubiquitin-like protein associating with human
RT            RAD51/RAD52 proteins.";
RL            GENOMICS 36:271-279(1996).
[4]
RN            SEQUENCE FROM N.A.
RP            SPECIES=HUMAN; MEDLINE: 97148692.
RX            MAHAJAN R., DELPHIN C., GUAN T., GERACE L., MELCHIOR F.;
RT            "A small ubiquitin-related polypeptide involved in targeting RanGAP1
RT            to nuclear pore complex protein RanBP2.";
RL            CELL 88:97-107(1997).
[5]
RN            SEQUENCE FROM N.A.
RP            SPECIES=HUMAN; MEDLINE: 97133418.
RX            MATUNIS M.J., COUTAVAS E., BLOBEL G.;
RT            "A novel ubiquitin-like modification modulates the partitioning of
RT            the Ran-GTPase-activating protein RanGAP1 between the cytosol and the
RT            nuclear pore complex.";
RL            J. CELL BIOL. 135:1457-1470(1996).
[6]
RN            SEQUENCE FROM N.A.
RP            SPECIES=HUMAN; TISSUE=PLACENTA;

```

```

RA MEDLINE, 97064180.
RA OKURA T., GONGE L., KAMITANI T., WADA T., OKURA I., WEI C.F.,
RA CHANG H.M., YEH E.T.H.
RT "Protection against Fas/APO-1- and tumor necrosis factor-mediated
RT cell death by a novel protein, sentrin."
RL J. IMMUNOL. 157:4277-4281(1996).
[7]
RN RN
RP SEQUENCE FROM N.A.
RC SPECIES-MOUSE; STRAIN-ICR;
RX MEDLINE; 98126440.
RX HOWE K., WILLIAMSON J., BODDY M.N., SHEER D., FREEMONT P.S.,
RA SOLOMON E.;
RT "The ubiquitin-homology gene PIC1: characterization of mouse (Pic1)
RT and human (UBI1) genes and pseudogenes."
RL GENOMICS 47:92-100(1998).
[8]
RN RN
RP STRUCTURE BY NMR.
RC SPECIES-HUMAN;
RX MEDLINE; 98319859.
RA BAYER P., ARNDT A., METZGER S., MAHAJAN R., MELCHIOR F., JAEHNKE R.,
RA BECKER J.;
RT "Structure determination of the small ubiquitin-related modifier
RT SUMO-1."
RL J. MOL. BIOL. 280:275-286(1998).
[9]
CC CC
CC -1- FUNCTION: ASSOCIATES WITH RAD51/RAD52. INVOLVED IN TARGETING
CC RANGAI TO THE NUCLEAR PORE COMPLEX PROTEIN RANBP2.
CC -1- SIMILARITY: TO UBIQUITIN. BELONGS TO THE SMT3 SUBFAMILY.
CC -----
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC CC
DR EMBL; X99586; E274603; -.
DR EMBL; U61397; G1518694; -.
DR EMBL; U38784; G1574948; -.
DR EMBL; U67122; G1762973; -.
DR EMBL; U72722; G1703503; -.
DR EMBL; U83117; G1769602; -.
DR EMBL; AF033353; G2645737; -.
DR MM; 601912; -.
DR PDB; 1A5R; 14-OCT-98.
DR PROSITE; PSS0053; UBIQUITIN_2; 1.
KW 3D-STRUCTURE.
FT DOMAIN 20 101 UBIQUITIN-LIKE.
SQ SEQUENCE 101 AA; 11557 MW; EC25C1FE CRC32;

Query Match 5.2%; Score 149; DB 1; Length 101;
Best Local Similarity 28.6%; Pred. No. 4,07e-06;
Matches 20; Conservative 23; Mismatches 26; Indels 1; Gaps 1;

Db 22 IKLVIGDSSEI-HFKYKMTTHLKIKESYCORQGVPMNSLRFEGORADNHTPKEL 80
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Oy 341 LRLRVQGEKHEQMELISPDSPPLKVLMSHYEAMGSLSGHRLSPFDQTKSLGKELPADL 400
Db 81 GMEEDVLEY 90
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Oy 401 GIESDLLEY 410

RESULT 6 STANDARD; PRT; 103 AA.
AC P55854;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE UBIQUITIN-LIKE PROTEIN SMT3A.
GN SMT3H1 OR SMT3A.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;

```


[illegible]

```

RN      [1] SEQUENCE FROM N.A.
RP      TISSUE-BRAIN.
RC      MEDLINE; 96080171.
RX      ZHANG Y., O'CONNOR J.P., SIONI M.C., SRINIVASAN S., DUTRA A.,
RA      NUSSBAUM R.L., DREYFUSS G.;
RT      "The fragile X mental retardation syndrome protein interacts with
RT      novel homologs FXR1 and FXR2."
RL      EMBO J. 14:5358-5366(1995).
CC      -1 FUNCTION: RNA-BINDING PROTEIN. INTERACTS WITH FXR1 AND FXR1.
CC      -1 SUBCELLULAR LOCATION: CYTOSOL;CYTOSOL.
CC      -1 SIMILARITY: BELONGS TO THE FXR1 FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; U31501; G1098637; -
DR      PFM; PFM00013; KH-domain; 2.
DR      HSSP; Q06787; 2EMR.
KW      RNA-BINDING.
FT      DOMAIN 232 261 KH.
FT      DOMAIN 295 324 KH.
FT      DOMAIN 414 418 POLY-SER.
FT      DOMAIN 544 552 POLY-ARG.
FT      DOMAIN 584 594 POLY-ARG.
SQ      SEQUENCE 673 AA; 74128 MW; 36F7ICF6 CRC32;

Query Match 4.0%; Score 113; DB 1; Length 673;
Best Local Similarity 32.4%; Pred. 1.93e-01;
Matches 23; Conservative 18; Mismatches 26; Indels 4; Gaps 4;

Db 471 RDPTRGESRRRPPTGGGRCGPPAPRPTSTKRNSSISYKDKPDSNPYSLDITSEP-EP 529
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 8 RGPFRSGRCGRGARAGARGAR-CPRARQSPARLIPDTVLVDLV-SDSDE-EVLEAVADPEV 64
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Db 530 PVDSEGEPEPP 540
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 65 PVARLPAPAKP 75

RESULT 11 STANDARD; PRT; 2142 AA.
ID BAT2_HUMAN
AC P46634;
DT 01-FEB-1996 (REL. 33, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE LARGE PROLINE-RICH PROTEIN BAT2 (HLA-B-ASSOCIATED TRANSCRIPT 2).
GN BAT2.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAEOA; CHORDATA; VERTERATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-T-CELL;
RX MEDLINE; 90192810.
RA BAKERJI J., SANDS J., STROMINGER J.L., SPIES T.;
RT "A gene pair from the human major histocompatibility complex encodes
RT large proline-rich proteins with multiple repeated motifs and a
RT single ubiquitin-like domain."
RL PROC. NATL. ACAD. SCI. U.S.A. 87:2374-2378(1990).
RN [2]
RP SEQUENCE OF 1-1860 FROM N.A.
RX MEDLINE; 93272029.
RA IRIS F.J.M., BOUGUERET L., PRIEUR S., CATERINA D., PRIMAS G.,
RA PERROT V., JURKA J., RODRIGUEZ-TOME P., CLAVERIE J.-M., DASSSET J.,
RA COHEN D.;
RT "Dense A1u clustering and a potential new member of the NF kappa B
RT family within a 90 kilobase HLA class III segment."

```

RL NAT. GENET. 3:137-145(1993).
 CC -1- FUNCTION: UNKNOWN.
 CC -1- TISSUE SPECIFICITY: LIMITED TO CELL-LINES OF LEUKEMIC ORIGIN.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M33509; G179339; -
 DR EMBL: M33518; G179345; -
 DR EMBL: M33512; G179345; JOINED.
 DR PIR: B35098; B35098.
 DR PIR: B36152; B36152.
 DR MIM: 142580; -
 KW REPEAT.
 FT DOMAIN 519 524 POLY-PRO.
 FT 636 657 GLN-RICH.
 FT DOMAIN 684 688 POLY-PRO.
 FT 699 704 POLY-PRO.
 FT DOMAIN 814 821 POLY-PRO.
 FT 1340 1345 POLY-GLY.
 FT DOMAIN 1398 1403 POLY-GLY.
 FT DOMAIN 1436 1442 POLY-PRO.
 FT 1982 1991 POLY-PRO.
 FT DOMAIN 41 1795 4 X 57 AA TYPE A REPEATS.
 FT REPEAT 41 95 1-1.
 FT REPEAT 98 154 1-2.
 FT REPEAT 281 337 1-3.
 FT 1740 1795 1-4.
 FT DOMAIN 337 549 2 X TYPE B REPEATS.
 FT REPEAT 337 418 2-1.
 FT REPEAT 476 549 2-2.
 FT DOMAIN 1899 2089 3 X 50 AA TYPE C REPEATS.
 FT REPEAT 1899 1948 3-1.
 FT REPEAT 1965 2014 3-2.
 FT REPEAT 2040 2089 3-3.
 FT CONFLICT 57 57 R -> A (IN REF. 2).
 FT CONFLICT 109 109 Q -> S (IN REF. 2).
 FT CONFLICT 414 414 P -> PPRKPGAGNNGPP (IN REF. 2).
 FT CONFLICT 532 532 T -> K (IN REF. 2).
 FT CONFLICT 682 682 Q -> K (IN REF. 2).
 FT CONFLICT 730 730 E -> D (IN REF. 2).
 FT CONFLICT 750 750 L -> R (IN REF. 2).
 FT CONFLICT 834 834 A -> T (IN REF. 2).
 FT CONFLICT 1035 1035 G -> A (IN REF. 2).
 FT CONFLICT 1068 1068 M -> L (IN REF. 2).
 FT CONFLICT 1285 1285 P -> R (IN REF. 2).
 FT CONFLICT 1400 1400 G -> A (IN REF. 2).
 FT CONFLICT 1611 1611 T -> S (IN REF. 2).
 FT CONFLICT 1729 1729 G -> A (IN REF. 2).
 SQ SEQUENCE 2142 AA; 227840 MW; 2CFEF88A CRC32;

Query Match 4.0%; Score 113; DB 1; Length 2142;
 Best Local Similarity 29.5%; Pred. No. 1.93e-01;
 Matches 31; Conservative 30; Mismatches 38; Indels 6; Gaps 6;

Db 1012 KRGCEYFARG-RGFRGTYGRRGGOANSATVESEFEMMGVE-VQGGDTTLLEPEAA-M 1068
 QY 6 RRGGP-RRSGGARRRARRRARRRSPARLIPDTLVLDVDSDEVLADPVEV 64
 Db 1069 P-ARHGARVQSMRKSPGAGGAKOAKAPMVRVIMLQTRRLPHR 1112
 QY 65 PVARLPAPAKPEQDSDSEGAEGPAGPRLTVRR-RRRRLDP 108

RESULT 12
 ID ISL2 BRARE STANDARD: PRT: 359 AA.
 AC P53406;

DT 01-OCT-1996 (REL. 34, CREATED)
 DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
 DE INSULIN GENE ENHANCER PROTEIN ISL-2 (ISLET-2).
 GN ISL2 OR ISL-2.
 OS BRACHYDANTO ZEBRAFISH (ZEBRA DANIO).
 CC ENKAROTA; METACOA; CHORDATA; VERTEBRATA; ACTINOPTERYGII; NEOPTERYGII;
 CC TELEOSTEI; ETELEOSTEI; OSTARIOPTYSII; CYPRINIFORMES; CYPRINOIDEA;
 CC CYPRINIDE; RASBORINAE; DANIO.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-EMBRYO;
 RX MEDLINE; 96005022.
 RA TOKUMOTO M., GONG Z., TSUBOKAWA T., HEW C.L., UYEMURA K., HOTTA Y.,
 RA OKAMOTO H.;
 RT "Molecular heterogeneity among primary motoneurons and within
 RT myotomes revealed by the differential mRNA expression of novel
 RT islet-1 homologs in embryonic zebrafish."
 RL DEV. BIOL. 171:578-589(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 96125154.
 RA APPEL B., KORZ V., GLASGOW E., THOR S., EDLUND T., DAVID L.,
 RA EISEN J.;
 RT "Motoneuron fate specification revealed by patterned LIM homeobox
 RT gene expression in embryonic zebrafish."
 RL DEVELOPMENT 121:4117-4125(1995).
 CC -1- FUNCTION: BINDS TO ONE OF THE CIS-ACTING DOMAIN OF THE INSULIN
 CC GENE ENHANCER. MAY BE INVOLVED IN SUBTYPE SPECIALIZATION OF
 CC PRIMARY MOTONEURONS.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC -1- DEVELOPMENTAL STAGE: FIRST EXPRESSED AFTER 15 H OF GASTRULATION,
 CC SEGMENTALLY IN THE VENTRAL REGION OF THE SPINAL CORD AND THEN
 CC LATER IN THE DORSAL REGIONS (ROHAN-BEARD NEURONS).
 CC -1- SIMILARITY: TO OTHER HOMEBOX DOMAINS. BELONGS TO THE LIM
 CC SUBFAMILY.
 CC -1- SIMILARITY: CONTAINS 2 LIM DOMAINS. THE LIM DOMAIN BINDS 2 ZINC
 CC IONS.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: D38453; G1037166; -
 DR EMBL: U09403; G487805; -
 DR EMBL: X88803; G871001; -
 DR PROSITE; PS00027; HOMEBOX_1; 1.
 DR PROSITE; PS00478; LIM_DOMAIN_1; 2.
 DR PROSITE; PS50023; LIM_DOMAIN_2; 2.
 DR PROSITE; PS50071; HOMEBOX_2; 1.
 DR PFM; PF00045; homeobox; 1.
 DR PFM; PF00412; LIM; 2.
 DR HSSP; P02836; 1HDD.
 KW HOMEBOX; DNA-BINDING; DEVELOPMENTAL PROTEIN; NUCLEAR PROTEIN;
 KW REPEAT; LIM MOTIF; METAL-BINDING; ZINC; MULTIGENE FAMILY.
 FT DOMAIN 27 80 LIM.
 FT 143 143 HOMEBOX.
 FT DNA_BIND 191 250 GLN-RICH.
 FT DOMAIN 255 359
 SQ SEQUENCE 359 AA; 40227 MW; DC592CD0 CRC32;

Query Match 3.9%; Score 111; DB 1; Length 359;
 Best Local Similarity 30.3%; Pred. No. 3.33e-01;
 Matches 23; Conservative 23; Mismatches 24; Indels 6; Gaps 6;

Db 174 PVROPFRHHVHKOSEKTRTVLNEKOLHTLRTCYVANRPD-ALM-KE-QLVEMTG- 229
 QY 192 PLPOSSSRKSRKHTLAKLREV-N-KRLDLSCLSPKQHSPPALDSTDEYVLVSGP 249

[illegible]

```

CC RNA DEPENDENT APASE ACTIVITY, AND RELEASE OF THE MRNA FROM THE
CC DNA TEMPLATE (BY SIMILARITY).
CC -1- SUBUNT: HOMOHXAMER (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RNP).
CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-43 IS THE INITIATOR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-slb.ch/announce/
CC or send an email to license@isb-slb.ch).
CC -----
CC
CC DR EMBL; X95444; E220353; -.
CC DR EMBL; X95444; E220354; ALT_INIT.
CC DR HSSP; P03002; 1A62.
CC KW TRANSCRIPTION TERMINATION; HELICASE; ATP-BINDING; RNA-BINDING.
CC FT DOMAIN 38 43 RNA-BINDING (RNP2) (BY SIMILARITY).
CC FT DOMAIN 343 346 RNA-BINDING (RNP1) (BY SIMILARITY).
CC FT NP_BIND 458 465 ATP (POTENTIAL).
CC FT DOMAIN 33 36 POLY-ARG.
CC FT DOMAIN 168 172 POLY-ARG.
CC FT DOMAIN 280 283 POLY-GLY.
CC SQ SEQUENCE 707 AA; 76544 MW; BC13166 CRC32;
CC
CC Query Match 3.9%; Score 112; DB 1; Length 707;
CC Best Local Similarity 48.5%; Pred. No. 2,53e+01;
CC Matches 16; Conservative 3; Mismatches 13; Indels 1; Gaps 1;
CC
CC Db 230 RDGGRDGRDGRDGRDGRDGRDGRDGRGGRDRNKK 262
CC | | | | | | | | | | | | | | | | | | | |
CC Qy 6 RGRGRSRGGRGARRRARG-RGPRARQSPAR 37
CC
CC RESULT 15
CC ID FBPI_DROME STANDARD; PRT; 1030 AA.
CC AC Q04691;
CC DT 01-JUN-1994 (REL. 29, CREATED)
CC DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
CC DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
CC DE FAT-BODY PROTEIN-1 PRECURSOR (P1 PROTEIN).
CC GN FBPI OR P1.
CC OS DROSOPHILA MELANOGASTER (FRUIT FLY).
CC OC EURARCTOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
CC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYROIDEA;
CC CC DROSOPHILIDAE; DROSOPHILA.
CC [1]
CC RP SEQUENCE FROM N.A.
CC RC STRAIN-CANTON-S; TISSUE=FAT BODY;
CC RX MEDLINE; 90339481.
CC RA MASCHAF F., DUBERTRET M.-L., THEONOD P., CLAVERIE J.-M.,
CC LEPESSANT J.-A.,
CC "Structure of the ecdysone-inducible P1 gene of Drosophila
CC melanogaster.";
CC J. MOL. BIOL. 214:359-372(1990).
CC [2]
CC RP SEQUENCE FROM N.A.
CC RC STRAIN-CANTON-S;
CC RX MEDLINE; 94123973.
CC RA LARIE P., NASR F., LEPESSANT J.-A., DEUTSCH J.;
CC "Deletion scanning of the regulatory sequences of the Fbpi gene of
CC Drosophila melanogaster using P transposase-induced deficiencies.";
CC GENETICS 135:801-816(1993).
CC CC -1- TISSUE SPECIFICITY: FAT BODY.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED DURING THE LATE THIRD LARVAL
CC STAGE.
CC -1- INDUCTION: BY ECDYSONE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: X54997; G8320; -
 DR EMBL: X69965; G7961; -
 DR PIR: S11034; S11034; -
 DR FLYBASE: FBgn0000639; Fbpl.
 DR PFAM: PF00372; hemocyanin; 2.
 KM SIGNAL: GLYCOPROTEIN.
 FT SIGNAL 1 16 POTENTIAL.
 FT CHAIN 17 1030 FAT-BODY PROTEIN-1.
 FT DOMAIN 360 371 POLY-ASP.
 FT DOMAIN 580 590 POLY-ASP.
 FT CARBOHYD 741 741 POTENTIAL.
 SO SEQUENCE 1030 AA; 119365 MW; A3AC5F31 CRC32;

Query Match 3.9% Score 111; DB 1; Length 1030;
 Best Local Similarity 20.9% Pred. No. 3.33e-01;
 Matches 42; Conservative 66; Mismatches 76; Indels 17; Gaps 15;

Db 559 TDDEILEMQLQRRROQLQK-OND-DD-D-NDDVNVVHROGLRSRRSLPRLRQNNRL 615
 QY 156 SEDDALPS-GSPWRKKLKKCKEKKMEFPDQISPLQPSSRNKSRKHTALQKLR 214
 Db 616 SEIVLHNRQ-LVARLNOESIAQGLIEQQQLINPRLTOSERYALRLNOIRINSQSR 674
 QY 215 VNKRLQDLRSCLSPKOHSPALQS--TDDEVVLVEGPLYPOSRL-FTLK-ICRADLYR 270
 Db 675 QVLAQIGIEQRIQIEGOVLSOVVNSLRGOVIDQROVESLIADVLGRLGQVIMTII 734
 QY 271 --LP-V-RMSEPLQVVDHMANHLSVSPNRIILLFG-ESE-LSPATPSTL-KLSVADII 323
 Db 735 RQVVDNNSQIDRNGLGIRL 755
 QY 324 DCVVLASSSEATE-TSQELRL 343

Search completed: Fri Feb 18 11:59:27 2000
 Job time : 51 secs.

Db 301 FGESELSPTATPSTLKLGVADIIIDCVIASSSEATFETISOELRLRVQGEKQHOMLEISISP 360
QY 301 FGESELSPTATPSTLKLGVADIIIDCVIASSSEATFETISOELRLRVQGEKQHOMLEISISP 360
Db 361 DSPILKMSHHEFAMGLSGHLSFFPDGTLKSGKELPADLGESGDLIEVWG 412
QY 361 DSPILKMSHHEFAMGLSGHLSFFPDGTLKSGKELPADLGESGDLIEVWG 412

RESULT 2
ID 057686 PRELIMINARY; PRT: 102 AA.
AC 057686;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SUMO-1 PROTEIN.
GN SUMO-1.
OS XENOPUS LAEVIS (AFRICAN CLAWED FROG).
OC EDUARIOTA, METAFOA; CHORDATA; VERTEBRATA; AMPHIBIA; BATRACHIA; ANURA;
OC MESOBATRACHIA; PIPOIDEA; PIPOIDEA; XENODODINAE; XENOPUS.
RN [1]
RX SEQUENCE FROM N.A.
RP MEDLINE: 98089176.
RA SATOH H., SPARROW D.B., SHIOMI T., PU R.T., NISHIMOTO T., MOHUN T.J.,
RA DASSO M.,
RT "Dc9p and the conjugation of SUMO-1 to Rangap1 and Ranbp2.";
RL CURR. BIOL. 8:121-124(1998).
DR EMBL: 297073; E323533;
SQ SEQUENCE 102 AA; 11703 MW; F3622FE3 CRC32;

Query Match 5.6%; Score 161; DB 13; Length 102;
Best Local Similarity 26.6%; Pred. No. 4.33e-07;
Matches 21; Conservative 27; Mismatches 30; Indels 1; Gaps 1;

Db 14 GDRKDGDPYIKLYIGODSSEI-HPKVMTKTHLKKKESYRQGVPMNRLPFEQORI 72
QY 332 SEATFSEELRLRVQGEKQHOMLEISISPSPILKMSHHEFAMGLSGHLSFFDGTKL 391
QY 392 SGKELPADLGESGDLIEV 410

Db 73 SDHQTPRELMEEDVIEV 91
QY 392 SGKELPADLGESGDLIEV 410

RESULT 3
ID 023759 PRELIMINARY; PRT: 115 AA.
AC 023759;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE UBIQUITIN PROTEIN.
OC CICER ARIETINUM (CHICKPEA) (GARBANZO).
OC EUKARYOTA; VIRIDIPHYTA; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
OC FABALES; FABACEAE; PAPILIONOIDEAE; CICER.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. CASTELLANA; TISSUE=ETIOLATED EPICOTYLS;
RA MUNOZ F.J., DOPICO B., LABRADOR E.,
RL PLANT PHYSIOL. 116:1605-1605(1998).
DR EMBL: AJ001901; E354254;
SQ SEQUENCE 115 AA; 12858 MW; 89C22D7C CRC32;

Query Match 5.6%; Score 161; DB 10; Length 115;
Best Local Similarity 25.0%; Pred. No. 4.33e-07;
Matches 19; Conservative 30; Mismatches 25; Indels 2; Gaps 2;

Db 18 PNDAAH-INLKVAGDGS-FFIKRTQTKLIMATCDROSVDLSIAFLFGRLRA 75
QY 334 ATETSQELRLRVQGEKQHOMLEISISPSPILKMSHHEFAMGLSGHLSFFDGTKL 393
Db 76 EOTPELDMEDGDEID 91
QY 394 KELPADLGESGDLIEV 409

RESULT 4
ID 057114 PRELIMINARY; PRT: 239 AA.
AC 057114;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE NONSTRUCTURAL PROTEIN P125-2 (FRAGMENT).
GN P125.
OS PESTIVIRUS TYPE 1.
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES. NO DNA STAGE; FLAVIVIRIDAE;
OC PESTIVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SNC;
RA OT F. BERRY E.S.;
RL SUBMITTED (FEB-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: U89438; G2707597;
KM NONSTRUCTURAL PROTEIN.
FT NON TER 1
FT NON TER 239
SQ SEQUENCE 239 AA; 26881 MW; 984208C6 CRC32;

Query Match 5.6%; Score 161; DB 14; Length 239;
Best Local Similarity 32.9%; Pred. No. 4.33e-07;
Matches 25; Conservative 19; Mismatches 31; Indels 1; Gaps 1;

Db 68 TENNDHNLNVAAGDGS-VVQFKIKRTPLSKLMKACERQGSMDRIIRFQSGPINET 126
QY 335 TETSQELRLRVQGEKQHOMLEISISPSPILKMSHHEFAMGLSGHLSFFDGTKL 394
QY 395 ELFPADLGESGDLIEV 410

Db 127 DTPAQLEMEDEDITDV 142
QY 395 ELFPADLGESGDLIEV 410

RESULT 5
ID 074186 PRELIMINARY; PRT: 117 AA.
AC 074186;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE PMT3P.
GN PMT3.
OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIASCOMYCETES;
OC SCHIZOSACCHAROMYCETALES; SCHIZOSACCHAROMYCETACEAE;
OC SCHIZOSACCHAROMYCETES.
RN [1]
RP SEQUENCE FROM N.A.
RA NISHIDE J., NISHI T., MURAKAMI Y., MATSUDA H., KAWAMUKAI M.,
RA TANAKA K.;
RT "The fission yeast ubiquitin-like modifier."
RL SUBMITTED (AUG-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: AB017187; D1033560;
SQ SEQUENCE 117 AA; 12934 MW; 9807EC2C CRC32;

Query Match 4.6%; Score 131; DB 3; Length 117;
Best Local Similarity 25.9%; Pred. No. 3.41e-03;
Matches 21; Conservative 24; Mismatches 35; Indels 1; Gaps 1;

Db 24 TSOQDVNPSTEHLNKLKVVAGDGNNEVF-FKIKTTFSKLKIKYCARQKSMNSIRFLVDG 82
QY 329 ASSEATFSEELRLRVQGEKQHOMLEISISPSPILKMSHHEFAMGLSGHLSFFDGT 388
Db 83 ERIRPDQTPALDMEDGDEID 103
QY 389 TKLSGKELPADLGESGDLIEV 409

RESULT 6
ID 063624 PRELIMINARY; PRT: 1173 AA.


```

SO      HOHO SAPIENS (HUMAN).
OC      EURARCTOA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
CC      CATARRHINI; HOMINIDAE; HOMO.
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=SKLELETAL MUSCLE;
RA      OSSORNE L.R., JUDAYEL D.M., COIGNET L.J., ZANI V.J., TSUI L.C.,
RL      SCHERER S.W., DYER M.J.;
RD      SUBMITTED (FEB-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
SO      EURL: AJ223980, F1249849.
DR      SEQUENCE 217 AA; 23365 MW; D15B84BE CRC32;

Query Match
Best Local Similarity 26.1%; Score 118; DB 4; Length 217;
Matches 40; Conservative 41; Mismatches 60; Indels 12; Gaps 11,

Db      59 RRRAGCGAEBRG-RER--RGRGASPRGGGLILLDLNDENSGNSFH-SSE-SLPKSTEP 113
          |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
11 RSRGGRGARRRARGARGCRPPAROSAPRIIPVTVLVDSDSEDEVLADVEVPVVRLP 70

Db      114 SPGCPQCSRPSPSPAPEGPEVEEAPRLOGERPDGIRTAGSTDEPMLTKKEPVPELL 173
          ::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Oy      71 APA-KPEDSDSDSGCAEG-PAGA-PTLVRRRRRRRLDGEAPVYPVS-GK-VGSSTL 125

Db      174 EAFAEATPFVEFPVPVEEAAGGTEDSGCAP 206
          : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy      126 NL-IPDNSSLKLCPSEPED-EADLTNKGSSPS 156

RESULT  11 PRELIMINARY; PRT; 425 AA.
ID      075038.
AC      075038.
DT      01-NOV-1998 (TREMBLREL. 08, CREATED)
DT      01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT      01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE      KIA00450 PROTEIN.
GN      KIA00450.
OS      HOMO SAPIENS (HUMAN).
OC      EURARCTOA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
CC      CATARRHINI; HOMINIDAE; HOMO.
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=BRAIN;
RX      MEDLINE: 98116662.
RA      SEIT N., OHITA M., NAGASE T., ISHIKAWA K., MIYAJIMA N., NAKAJIMA D.,
        NOMURA N., OHARA O.;
        "Characterization of cDNA clones in size-fractionated cDNA libraries
        from human brain.";
        DNA RES. 4:345-349(1997).
RL      DNA RES. 4:345-349(1997).
RD      EMBL: AB007919; D1033257; -.
SO      SEQUENCE 425 AA; 43842 MW; 0064808E CRC32;

Query Match
Best Local Similarity 4.1%; Score 117; DB 4; Length 425;
Matches 35; Conservative 25; Mismatches 43; Indels 8; Gaps 7,

Db      127 GR-PAAPSVSQRILRRASAPTQSQRGGRGFE--LY-L-GRTRDGSKGVADV-VPP 180
          |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Oy      7 GGSPRSKRGGRARRARGRGCRPROSPALPIPTVLVDLSVDSDEVELFAADPEVPV 66

Db      181 GPGPAPPAQGPQSGSPGCKAPAFAVESKPVVRPRPRVLGDGPGAPMAA 231
          : |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Oy      67 ARLPARPAPREDSDSDSGAEGPAG-APLTLVRRRRRLLD-PGEAPVPV 115

RESULT  12 PRELIMINARY; PRT; 771 AA.
ID      006359.
AC      006359.
DT      01-JUL-1997 (TREMBLREL. 04, CREATED)
DT      01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)
DT      01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE      HYPOTHEICAL 81.4 KD PROTEIN.
        MTCYI5C10.03C.

```

[illegible]

```

RESULT      14
ID           020393
AC           020393:
DT           01-NOV-1996 (TREMBLREL. 01, CREATED)
DT           01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT           01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE           F44D12.4 PROTEIN.
GN           F44D12.4.
OS           CAENORHABDITIS ELEGANS.
OC           EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC           RHABDITINA; RHABDITOIDEA; RHABDITIINAE; PELODERINAE; CAENORHABDITIS.
LN           [1]
RN           SEQUENCE FROM N.A.
RP           COLES L.;
RL           SUBMITTED (DEC-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
TX           [2]
RP           SEQUENCE FROM N.A.
RL           MEDLINE; 94150718.
TX           WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BEKKS M.,
RA           BONFIELD J., BUTTON J., CONNELL M., COPEST T., COOPER J., COULSON A.,
RA           CRAYTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA           GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIRI M., JOHNSTON L.,
RA           JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LARREILLE P.,
RA           LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA           PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOMKKEEN R.,
RA           SMALDON N., SMITH A., SONNHAMER E., STADEN R., SULSTON J.,
RA           THERIY-WIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA           WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLIDMAN P.;
RT           "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT           elegans";
RL           NATURE 358:32-38(1994).
DR           EMBL; Z68296; E1346770; -
SQ           SEQUENCE 352 AA; 39367 MW; 2EA853D3 CRC32;

```

THIS PAGE BLANK (USPTO)